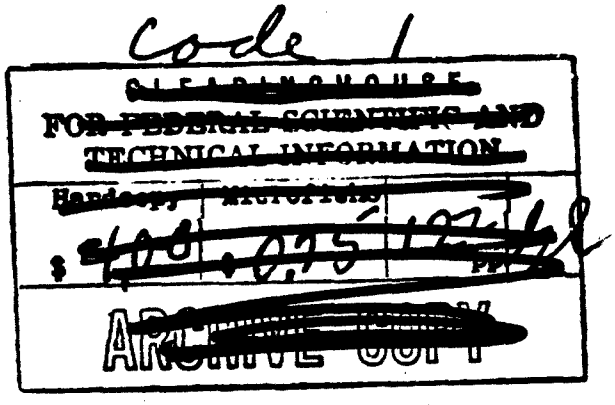
SELECTIVE DISSEMINATION OF INFORMATION: REVIEW OF SELECTED SYSTEMS AND A DESIGN FOR ARMY TECHNICAL LIBRARIES

William A. Bivona and Edward J. Goldblum



August 1966

Prepared for

U. S. ARMY NATICK LABORATORIES
Natick, Massachusetts

Contract DA 19 129-AMC-957 (N)

by

INFORMATION DYNAMICS CORPORATION
80 Main Street
Reading, Massachusetts

IDC 8074

SELECTIVE DISSEMINATION OF INFORMATION: REVIEW OF SELECTED SYSTEMS AND A DESIGN FOR ARMY TECHNICAL LIBRARIES

William A. Bivona and Edward J. Goldblum

August 1966

Prepared for

U.S. ARMY NATICK LABORATORIES
Natick, Massachusetts

Contract #DA 19-129-AMC-957 (N)

by

INFORMATICS INCORPORATION
[illegible] Main Street
Reading, Massachusetts

TABLE OF CONTENTS

TABLE OF CONTENTS (cont'd)

TABLE OF CONTENTS (cont'd)

LIST OF TABLES AND FIGURES

ABSTRACT

This report presents an analysis of over eighteen small, intermediate, and large-scale systems for the selective dissemination of information (SDI). Systems are compared and analyzed with respect to design criteria and the following ten system parameters: information input, methods of indexing and abstracting, user interest profile construction, user group, profile/document match strategy, profile maintenance and updating, products disseminated, operating statistics accumulated, ADP equipment utilized, and results reported. From the results of the analysis, criteria are drawn for the design of an SDI system that is generally applicable to a broad range of Army technical libraries. The design criteria are used as a basis for designing an SDI system which has the following desired characteristics: broad applicability, acceptance of information in many different formats, ability to yield high relevancy for specific information requirements, and high recall for general information requirements. This system requires a minimum automatic data processing configuration and operating personnel. Operating costs are minimized by utilizing standard graphic reproduction techniques for reproducing multiple copies of SDI notices. The user profile/document match strategy incorporates both Boolean and weighted selection criteria.

FOREWORD

Selective dissemination of information (SDI) holds the prospect of being the most efficient method for transferring large masses of information from the point of origin to the ultimate destination with a minimum of manual intervention. The SDI concept represents a large-scale effort to give information services an active role in the research and development process. It is no wonder, then, that the concept has captivated the imagination of a wide variety of technical information activities ranging from the information support facilities of chemical and drug industries to the aerospace information support activities sponsored by NASA.

Army technical libraries are presently investigating the role of SDI in the development of technical support for Army research and development projects. The investigation and application of SDI techniques is progressing along thre different fronts. Individual libraries, such as those which support the Army Electronics Command at Fort Monmouth, New Jersey, and the Army Biological Laboratory at Fort Detrick, Maryland, are creating SDI systems to met immediate local requirements. The Army Technical Libraries Improvement Studies (ATLIS) Program is supporting two other approaches -- the development of an SDI system to operate as a central switching point for Army research and development information and a generalized SDI system designed to be applicable to any Army technical library which supports local information requirements.

This report represents the results of the first phase of an effort to develop an SDI system for local use within a wide range of Army technical communities. Of necessity, this system's design had to be elastic enough to accommodate the entire range of individual Army technical library capabilities in terms of access to automatic data processing equipment and staff size and in terms of the local information requirements of research and development personnel served by the library.

It is anticipated that this system design will be implemented in a supplemental effort using Army Natick Laboratories as a test bed for conducting a field test on a group of 25 participants and an information input of over 1000 technical reports, books, journal articles, and other information formats during a nine month pilot operation.

This final report was prepared for the ATLIS program under the direction of Mr. Robert Martin, Chief Librarian, Army Natick Laboratories, Natick Massachusetts. All work was performed by Information Dynamics Corporation, 80 Main Stret, Reading, Massachusetts under Contract DA19-129-AMC-957 (N).

ACKNOWLEDGMENTS

The authors express their gratitude to the following persons for their invaluable assistance in the conduct of this study and systems design: Mr. David P. Waite, President, Information Dynamics Corporation, for his critical evaluation of the system design and his astute guidance in the conduct of this study; Mrs. Rita E. Morgan, company librarian, for her thorough compilation of Appendices B and C (bibliographies of documentation pertinent to SDI) and the Matrix Summary of SDI Systems Characteristics in Chapter II.

I. INTRODUCTION

The procedure known as Selective Dissemination of Information (SDI) has evolved to the present state of sophistication from its simple origin as a courtesy service accorded to favored patrons of a library. Based upon the patron's description of his interests and a history of what materials he had requested previously, the librarian formulates a mental pattern of the patron's information requirements. As new material is acquired and processed, the librarian takes note of those items which might be of interest to the particular patron and brings them to his attention, when time permits. If the information is not of interest, the patron restates his interests in more explicit terms, cancels obsolete requests, and adds entirely new interests.

The SDI system of today is a formalization of the casually executed procedures which presently relies on the memory and willingness of the librarian. In most currently operating SDI systems, the process of matching the patron's interest profile against the subject description of documents is expedited by programmed, high-speed digital computer. Subject index terms applied to documents, along with descriptive cataloging information and abstracts, are recorded on punched cards and then transcribed onto magnetic tape. The patron's interest profile and address are similarly encoded, keypunched, and transcribed to tape. The output derived from matching the two magnetic tapes is usually the identification numbers, descriptive cataloging information, and full abstracts of the documents that match specific patron's interests. These outputs, addressed by the computer, are sent daily, weekly or monthly to the patron, depending on the current document accession rate and number of subscribers.

After reviewing the computer output, the patron indicates whether the information supplied was:

a. of no interest,

b. of interest, but the full document is not required, or

c. of interest and the document is required.

This information is returned to the librarian, who takes the appropriate action. If the information supplied was of no interest, those index terms which caused a match between the patron's interest profile and the document description are deleted from the profile, so that similar information will not be retrieved in the next dissemination cycle. If the patron so indicates, the profile is modified to include entirely new interests.

An SDI system should really be thought of as a sub-system of a wide range of library services. As such, the SDI system must be carefully integrated with other library operations. Under the Army Technical

Library Improvement Studies (ATLIS) Program, the Army is attempting to co ordinate efforts to improve the services provided by all Army technical l ries. The ATLIS Program encompasses the improvement of all aspects of li operations, such as subject indexing, descriptive cataloging, vocabulary trol, circulation procedures, and interlibrary cooperation. The develop of SDI procedures is one part of this comprehensive improvement study pro gram.

This report outlines the design for an SDI system which is capabl of servicing personnel having a broad spectrum of subject interests and e gaging in a wide variety of activities. The design can be implemented by any Army technical library having access to a small, medium, or large sca digital computer and any reproduction facility ranging from office duplic tor to offset printing press. The system accepts any input which can be characterized by a linear notation for mechanized comparison with user pr files and which can be graphically reproduced onto an EAM card format for distribution to the user. The input can contain photographs, chemical an mathematical formulae, charts and graphs, and engineering drawings, diagr or flow charts.

The design of the SDI system presented in this report is based on study of over twenty currently operating SDI systems, including two which are operating at Army installations - one at Ft. Monmouth, New Jersey, an the other at Ft. Detrick, Maryland. This SDI system design incorporates those features of the systems studied which could be extrapolated to meet the capabilities and requirements of the widest selection of Army technic libraries. Care was taken to design an SDI system which would not requir the use of a large-scale digital computer and large amounts of computer t

In Chapter II of this report, the SDI concept is carefully define to distinguish it from other library services. To reflect the latitude i the missions that exist among Army technical libraries, a complete outlin of optional objectives is presented within which the SDI system can oper Selection of individual options leads to systems configurations tailored the specific requirements of individual libraries. A comparative review seventeen SDI systems is presented in Chapter III, and summarized in a ma trix which presents important features of each system. Chapter IV is dev to an analysis of the facilities to which Army technical libraries have a cess to demonstrate the general applicability of the SDI system design pr sented in Chapter V. A complete bibliography of the published literature SDI systems and a detailed description of each system studied are present in the Appendices.

II. SDI SYSTEMS - DEFINITION AND PURPOSE

SDI systems are defined below by indicating those elements of the system which are essential, and by differentiating SDI systems from other similar library services. Within this definition, one can select many different input options to reflect specific purposes dictated by the overall mission of the library. A description of these options provides a basis for differentiating the systems analyzed in Chapter III and for selecting the design features for an Army SDI System presented in Chapter V.

A. DEFINITION OF SDI

1. Luhn's Original Definition

The phrase "Selective Dissemination of Information (SDI)" was coined by Hans Peter Luhn in a paper published in 1958 (see Appendix C - Reference No. 50). In this paper, Luhn discussed an entire range of business-related information retrieval problems to which he proposed an integrated solution. Part of the solution which Luhn proposed involved the selective dissemination of information, which he defined in a subsequent paper (see Appendix C - Reference No. 51) as "that service within an organization which concerns itself with a channeling of new items of information, from whatever source, to those points within the organization where the probability of usefulness, in connection with current work or interests, is high."

In the system envisioned by Luhn, current information extracted from journals, newspapers, technical reports and other publications would be indexed by terms selected from the documents themselves. User information requirements would be indexed in the terms of the user's own vocabulary and retained as standing requests for information. Each incoming document description would be matched against the bank of standing requests by means of EDP equipment, which would, in the process, prepare a list of all individuals who might be interested in that document. A token description of the document would then be printed and distributed to all persons on the list. The token description might consist of title, authors, source and abstract. In recognition of the imprecision of the matching process and the changing interests of the users, Luhn provided a feedback loop to allow the user to modify his interest profile on the basis of information previously received. Thus, the user would be able to control the total volume of information disseminated to him and also the percentage of information that would be relevant to his current work or interests.

Luhn's definition leaves room for disseminating items of information, either recently acquired or newly published, to either individuals or groups of individuals. It was Luhn's original contention that a profile of group interests would be as useful as a profile for each individual's interests and that newly acquired information might be just as useful to individuals as information recently published.

2. Similarity of SDI to Other Information Services

In many ways, the SDI systems which grew from Luhn's propos are similar to other information services, especially in terms of implemen ation. Indeed, many existing SDI systems utilize the same information se strategy as that used in "retrospective" information retrieval. Procedur used in supplying document copies are also similar to the document loan p cedures of conventional libraries.

Because of this similarity to SDI systems, some conventiona library services have been erroneously labeled "Selective Dissemination o Information." For example, several documents which purport to describe S systems in reality describe systems for distributing subject classified a cessions lists. Only those systems which conform to the intent of Luhn's definition of SDI will be discussed in this report, even though other library services may closely approximate that definition.

3. SDI Systems Definition

By observing the experience gained by others in implementin and operating SDI systems since 1959, it is possible to draw a definition of SDI which does not include all of the elements originally included by Luhn, without abrogating Luhn's original intentions. One of the elements which is noticeably lacking in many present systems is the document auto-indexing feature (index terms are selected directly from the texts of abstracts or titles on the basis of frequency of occurrence, word length, o gross grammatical structure), although this feature is still used by the sent IBM SDI system. Because of the proven utility of a controlled index vocabulary and its wide use in retrospective retrieval systems, the auto-indexing feature is not included as a prerequisite for SDI. In a control vocabulary system, a word or phrase selected from a thesaurus is assigned represent a concept, even though that word or phrase may not appear in th text.

In this report, an SDI system is defined as any system whic has the following essential features:

(a) Input to the system consists of

(1) any information (for dissemination) which can be characterized by a string of characters and which can be graphically reproduced;

(2) descriptions of users' information requirements;

(3) a list of the users' addresses; and

(4) feedback from the users which indicates the degree of relevancy of output received and provides a basis for improving relevancy of output.

(b) Processes performed by the system consist of:

(1) matching descriptions of users' information requirements against descriptions of the contents of input documents;

(2) selection of document descriptions which match users' information requirements;

(3) modification of users' profiles on the basis of feedback from the users; and

(4) addressing of outputs selected.

(c) Output consists of

(1) document descriptions selected and addressed in (b); and

(2) statistics which indicate the operational characteristics of the system.

This operational definition permits the use of auto-indexing or controlled indexing vocabulary, includes the dissemination of all information, regardless of date of publication or format, and in no way limits the mode for obtaining feedback.

B. SDI SYSTEMS DESIGN OPTIONS

The design of an SDI system is markedly influenced by the purpose for which the system is established. Ostensibly, the purpose of an SDI system is to channel new information to points within an organization where the probability of usefulness is high. The amplification of this definition, however, leads to diverse points of view, as exemplified by the operational definitions adopted by individual systems.

1. Newly Acquired Information vs. Newly Generated Information

The first point of contention is the definition of what is "new" or "current" information. Some systems have adopted the convention that any information recently acquired by the system is by definition "current information." Some systems, however, use the date of publication as a criterion for screening out older material. The dissemination of older material to individuals who know of its existence or are already familiar

with its contents would have an adverse effect on how the system is received. On the other hand, older material not readily accessible to most of the system's clients might have considerable informative value.

2. Announcement vs. Self-Sufficient Information Package

A more central issue is that of whether the system should merely indicate the existence of information, or whether the system should strive to provide self-sufficient information packages. A previous study (see Appendix C - Reference No. 66) has shown that titles alone are sufficiently indicative to allow the reader to identify pertinent documentation Thus, if the purpose of the system is merely to indicate what information is available, this purpose can just as well be accomplished by the distribution of titles only. In attempting to provide complete and self-sufficien information packages, some SDI systems disseminate a range of products varying from indicative abstracts all the way to complete source documents. I an abstract is sufficiently informative, it may satisfy the immediate information requirements of the recipient and thus obviate the need for orderin the full document.

3. Machine-Readable Input vs. Non-Machine-Readable Input

The design of an SDI system is often dictated by a pragmatic approach to providing information services at minimal cost. Some services have access to suitable information already in machine-readable form. If this information consists only of title, author, and journal source, and if this bank of machine-readable data is compatible with the majority of user needs, then it is logical to tailor the system's design to process this type of input and dispense with providing full abstracts in order to provide a maximum SDI capability at a very low cost.

4. Responsibility for Document Supply vs. No Responsibility

If the SDI system utilizes externally-produced machine-reada data, the very practical question is raised as to whether the system shoul assume the responsibility for obtaining copies of all documents announced. Even if the system does not assume the financial burden of acquisition, additional personnel might be required to process document orders.

The document-supply question has varied answers, even for those systems which announce only information on hand. Some systems have solved this problem by ordering multiple copies of all acquisitions and others have solved it by routing single copies or by duplicating copies as required. A fourth solution requires that all information be viewed at th library when more than one request is received for the same document. Observance of copyrights might also prevent copying full texts.

5. Internal Information vs. External Information

In formulating the design of an SDI system, one must also co sider whether the system will handle internally as well as externally pro-

duced information. In Luhn's original presentation of a business intelligence system, explicit provision was made for disseminating internally-produced documents; but due to the administrative structure in which the SDI system is imbedded, the system may not have, or may not desire to have, access to internally-produced documents for the purpose of dissemination. In large organizations where internal communications are limited by organizational, locational or administrative boundaries, there is some merit in shunting internal communications through an SDI system which routes the information on an impartial basis. On the other hand, in a close-knit group, there may be no need to supplement regular communications channels.

6. Single Format Input vs. Multiple Format Input

The decision as to what formats (i.e., journal articles, patents, technical reports, books, product literature, newspaper articles, letters and memoranda, and symposium or course announcements) will be accepted by the system is based mostly on practical considerations, such as available manpower and coverage by other information services. Due to limited resources, it may be impossible to accept journal articles which are not indexed and abstracted in the regular course of providing other retrieval services. Conversely, extra resources may be set aside specifically for the purpose of covering journals, because these are not covered by normal library services.

7. Servicing Single Activities vs. Multiple Activities

In developing a useful design, the SDI systems planner must also examine the kinds of activities in which the system's clientele is engaged. Administrative personnel might be most interested in obtaining information which reflects internal operations rather than very specific technical information needed for technical task decisions. Individuals engaged in research, on the other hand, would be most interested in technological detail. Each type of activity generates its own information requirements, if serviced by the SDI system. These requirements, in turn, may be satisfied by different kinds of input information.

8. Cooperative Effort vs. Independent Effort

As the number of SDI systems increases, the purpose of any single system may be dictated in part by the availability of service from other systems. Because of the present degree of subject specialization and the interrelatedness of subject specialties, it is becoming increasingly difficult for any one system to cover information generated outside its own specialty, even though this information may in fact be pertinent. For this reason, SDI systems organized along subject-specialty lines might profitably support each other by trading SDI services. Indeed, the decision as to what materials should be handled by the SDI system may eventually be determined in part by what materials are <u>not</u> handled by other systems.

A network of cooperating systems would also increase the yield efficiency of each of the participating systems. The yield efficiency is

defined as the ratio of the number of units disseminated to the total number of input units. If a system supporting biologically oriented personnel were to input documents on electronics, the yield efficiency would be very low because most of the documents would not be interesting to any of the system's clients. Thus the cost of each notification would be inordinately burdened by the cost of input.

9. Summary

The above discussion is given to indicate that the process of formulating the design of an SDI system is a complex decision-making process involving the selection of many parameters for the optimization of many different conditions. Before a detailed systems design can be initiated, the objectives must be clearly defined so that explicit choices can be made with regard to the following input options:

* Announcement vs. self-sufficient information package
* Newly acquired information vs. newly generated information
* Internal information vs. external information
* Machine-readable input vs. non-machine-readable input
* Single format input vs. multiple format input
* Cooperative effort vs. independent effort
* Responsibility for document supply vs. no responsibility
* Coverage of library acquisitions only vs. unlimited coverage
* Coverage of presently utilized sources vs. presently unused sources
* Servicing single activities vs. multiple activities

Choices among the above options will define the objectives to be met by the system and will thus guide the design of the ultimate system's configuration for any given installation.

The next chapter will describe the design choices made for currently operating systems and the systems configurations resulting from these choices.

III. COMPARATIVE REVIEW OF SDI SYSTEMS

As a precursor to the design of an SDI system applicable to a wide range of Army technical libraries, seventeen operating SDI systems were studied in detail. The results of the study are summarized in Table I - Matrix Summary of SDI Systems Characteristics.

This chapter reviews the current status of SDI system development and discusses the detailed operations of seventeen SDI systems in terms of:

* Data input
* Document/user profile match strategy
* Data output
* Cost
* Measures of systems efficiency
* Effect of SDI on the user
* Effect of the user on other information services

Data was collected from published literature, personal correspondence, face-to-face and telephone communications, and preprinted questionnaire (Figure A-1). The full data collected is summarized for each system in Appendix A. It is significant to note that cost data is not available for most of the systems studied, although costs can be deduced from an analysis of the operating parameters. In SDI systems which form an integral part of larger information systems, it is difficult to prorate costs between SDI and other information services offered. It is also significant that the evaluation criteria used by each system vary to the extent that no valid comparisons can be made on the basis of published statistics.

A. CURRENT STATUS OF SDI SYSTEMS DEVELOPMENT

Over the past three years, there has been a marked increase in the number of operating SDI systems. The primary reasons for this are an increased recognition by both Government and industry of the importance of information services and an increasing reliance on the use of electronic data processing equipment to implement these services. The general trend towards automation in documentation is also facilitated by the ready availability of large banks of machine-readable data created by previously established automated systems.

1. Original Reception to SDI

When Luhn first disclosed his concept of a system for the selective dissemination of information, the idea was received by the public with some skepticism. First of all, the application of computers to information handling problems had just begun. Secondly, Luhn linked his SDI concept to his earlier presentation of a statistical approach to mechanized encoding and searching of the literary information. The newness of electronic data processing technology and statistical approaches to the analysis of information combined to produce an initially unfavorable reaction towards the SDI concept.

MATRIX SUMMARY OF

SYSTEMS	DOCUMENTATION INPUT	VOCABULARY CONTROL	PROFILING METHOD
CHEMICAL ABSTRACTS SERVICE (Joint Experiment)	References from Chemical Titles Magnetic Tapes	--	Keywords and/or Word Stems; Authors and/or Coden
CHEMICALS AND PHOSPHATES, LTD.	Journal Articles, Books	A.I.Ch.E. Thesaurus	Up to 12 Coded Keywords per Profile
DOUGLAS AIRCRAFT COMPANY	Documents	Coded Dictionary 11,000 Terms	Descriptor Weights
DOW CHEMICAL COMPANY	Magnetic Tapes of Chemical Titles and Documents	Authority Lists, A.I.Ch.E. Thesaurus Using Links and Roles	Descriptor Weights with Indication of Hit Threshold
ELI LILLY & COMPANY	Magnetic Tapes of Chemical Titles from Chemical Abstracts Service	--	Keywords and/or Word Stems; and Authors
IBM - CHICAGO	Articles from 150 Periodical Publications	200 IBM Identifiers, 600 Word Exclusion List, 2200 Word Inclusion List, Automatic Indexing	Weighted Terms w/values of -9 to +9 and/or Word Roots
IBM - OWEGO	Documents	DDC Thesaurus and List of "Open End" Terms with Reference Code	Descriptors and "Open End" Terms
IBM - POUGHKEEPSIE	IBM Documents, Patents, Minutes, DDC and NASA Documents, Books, Misc. Reports	A.I.Ch.E. Thesaurus supplemented by 2000 IBM Terms	Keywords from A.I.Ch.E. Thesaurus and IBM Supplemental Terms

ITEMS CHARACTERISTICS

C OF / / SES	PROFILE/ DOCUMENT MATCH	MATCH NOTICE	USER FEEDBACK NOTICE	% NOTICES a. RETURNED b. RELEVANT c. REQUEST DOCUMENT		SYS HARD
-	Predetermined Relevance of Profile/Chemical Titles Keywords	--	--		--	IBM
000	Single Keyword Match of Document/Profile Keyword	Title, Author, Source, Serial Number and Reference	*See Legend #2		--	IBM w 4K
525	Match of Document/Profile Weighted Terms	Library Index Card Coupled to Response Card	*See Legend #2 Abstract Deleted	a. b. c.	-- 50-60% --	IBM IBM
-	Document Assigned Hit Value = Sum of Weights of Profile Terms Matched	--	--		--	-
343	Boolean Inclusion, Exclusion, and Negation	*See Legend #1	*See Legend #2		--	IBM 140 Memory 705, 40 ory: 60 Printer Tape Dr
ers files	Formula of Sum of Identical Words = Predetermined Profile Hit Level	*See Legend #1	*See Legend #2	a. b. c.	-- 66% 11%	IBM 140 4 Tape
000	% Profile/Document Match Terms on Complete Words or Phrases	*See Legend #1	*See Legend #2	a. b. c.	73% 58% 24%	IBM IBM
site	% Profile/Document Match and MUST, NOT and IMPERATIVE Designations	Bibliographic Entry, Abstract, Index Terms in Order of Probable Relevance	PORT-A-PUNCH® Provides for Optional Responses	a. b. c.	-- 75-80% --	IBM

MATRIX SUMMARY OF

SYSTEMS	DOCUMENTATION INPUT	VOCABULARY CONTROL	PROFILING METHOD
IBM - YORKTOWN HEIGHTS	Documents	Controlled Vocabulary of Single and Multiple Word Descript[illegible]	Descriptive Words and Phrases [illegible] 120 Charac-ters w/'MUST MAY, NOT [illegible]
[illegible]TUTE FOR SCIENTIFIC INF[illegible]MATION	1700 Journals All U.S. Patents	Bibliograph[illegible] References Serve as Am-biguous Index-ing Terms	[illegible] Questions [illegible]
[illegible] UNIVERSI[illegible]	All Pertinent [illegible]	None	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible]	[illegible]	[illegible]	[illegible]
[illegible] LABORATORY	[illegible]	[illegible]	[illegible] MESH Tapes

S CHARACTERISTICS (CONT'D)

R OF S/ LES	PROFILE/ DOCUMENT MATCH	MATCH NOTICE	USER FEEDBACK NOTICE	% NOTICES a. RETURNED b. RELEVANT c. REQUEST DOCUMENT	H
--	Fixed Number of Terms Complying with MUST, MAY, NOT Constraints	*See Legend #1	*See Legend #2	--	I
--	Exact Match	Original Question and Citations Distributed Weekly	Reverse Side Has Provision for Hard Copy Order	--	
[illegible]	Probability Formula or Significance [illegible] Word Clusters	Summary of Document Content and [illegible] [illegible] Option	Automatic Increment [illegible] Adjustment [illegible] Significance Value of Terms	[illegible]	[illegible]
[illegible]	Any Document/ Profile Descriptor Match	Printout of [illegible] Entry and User's Name, Location, etc.	Codaphone Accepts User Feedback	5 - [illegible] Document Requests per Month	[illegible]
[illegible]	[illegible]	*See Legend #1 [illegible] Accession [illegible] and Price of Document	*See Legend #[illegible]	[illegible]	[illegible]
Groups	[illegible] the MATCH Program	[illegible] Abstract or Document at User's Option	*See Legend #2	--	[illegible]
[illegible]	[illegible] Terms [illegible] Boolean Expressions in Profile	Printout of All Match Citations	*See Legend #2	a. [illegible]-73% b. [illegible]-73% c. 25%	[illegible]

SYSTEMS	DOCUMENTATION INPUT	VOCABULARY CONTROL	PROFILING METHOD
U.S. ARMY ELECTRONICS COMMAND	Printout of DDC TAB Abstracts	DDC Thesaurus	Selected Te from DDC Th saurus *Descriptor Groups with Descriptors Section
U.S. DEPARTMENT OF INTERIOR BUREAU OF RECLAMATION	Documents Preprints Translations	Thesaurus Wordlist 3260 Terms	Limit of 20 Descriptors 4 May be Weighted by Asterisks

* LEGEND:

1. Complete bibliographic entry including document numb of user; and various optional instructions to user p

2. Notice contains provision for user to indicate degre request changes in his profile, change of address ar

3. Four experimental match runs using various types of
   a. Chemical Titles yielded 55% notifications of int
   b. Also, Chemical Titles yielded 65%
   c. Sandia Documents produced 59%
   d. N.S.A. Keyworded Abstracts gave an average of 5

4. In this system, users are instructed to return only

BA... OF EP... / FI... ES	PROFILE/ DOCUMENT MATCH	MATCH NOTICE	USER FEEDBACK NOTICE	% NOTICES a. RETURNED b. RELEVANT c. REQUEST DOCUMENT	S... HA...
...)	Match of Asterisked Descriptor and Negative Descriptors	Printout of Pertinent TAB Abstracts in Booklet Form from TAB Tapes	Sheet Provided for Listing of Required Documents by "AD" Number	a. -- b. 70% c. 15%	BU... B...
...Users	One Weighted Term Match and/or Three Unweighted Terms	*See Legend #1	*See Legend #2	a. 47% b. *See Legend #4 c. 11%	Flexo... HONEY... HONEY...

...rac... and keywords; name, address, and location
... fo... feedback.

...eva...ce, request document, make pertinent comments,
...rti...ent comments regarding the SDI system.

...re ...ade. Results are as follows:

...cat...ons of interest.

... re...evant notices.

B

Shortly after publication of his original paper in 1958, however, Luhn implemented the first automated SDI system on an IBM 650 general-purpose digital computer located at Yorktown Heights, New York. The encouraging results obtained from this early attempt led to the development of a completely documented systems package called "SDI-2." As a direct result of making this documented system available, Luhn fostered the development of six other SDI systems during 1961-2. Three of these were installed at IBM facilities and three at individual manufacturing companies. During this same period, SDI-3 and SDI-4 were developed.

2. Characteristics of SDI Development

Due to the newness of the SDI concept, the development of early SDI systems was evolutionary in nature. The experience gained from initial SDI installations was used to establish design criteria for subsequent systems. Even within a single installation, an SDI system underwent a process of evolution guided by initial results. An extreme example of such evolution was IBM's consolidation of four SDI systems into a single system, located at the Thomas J. Watson Research Center at Yorktown Heights. The consolidated system includes some of the features of the systems formerly located at Yorktown Heights, Poughkeepsie, Chicago and Owego.

The state-of-the-art of SDI systems development can be characterized as still being open to further innovation, study, and evaluation, even though some systems have been in continuous operation for over five years. An SDI system is still far from being an off-the-shelf item which operates in a completely predictable way and which can be assimilated into another environment without modification. Although there are over a dozen systems operating successfully today, the reasons for successful results are not clearly understood in terms of the significant differences between them. A complete understanding of all the mechanisms that operate within an SDI system must await a comparative analysis under carefully controlled conditions in which only one of the many recognizable operating parameters is varied at a time.

Because present systems are constantly undergoing revision, systems documentation is not readily available and is usually out of date when published in the open literature. For this reason, it was necessary to contact the directors of individual systems to obtain the latest status of each system.

3. Summary and Conclusions

In summary, development of SDI systems is still undergoing a period of rapid growth in which many innovations can be expected. However, with the increasing availability of large data banks in machine-readable form, recognizable clusters of SDI systems are emerging. For the most part, these are centered around machine-readable products obtainable from Defense Documentation Center, National Aeronautics and Space Administration, National Library of Medicine, and Chemical Abstracts Service. Since these organizations announce a sizeable fraction of the nation's current literature (in

such publications as the DDC Technical Abstract Bulletin (TAB), the NASA Scientific and Technical Aerospace Reports (STAR), the NLM Index Medicus, and the CAS Chemical Titles and Chemical-Biological Activities), and since these announcement journals are published periodically, the utilization of the machine-readable by-products from which they are generated is extremely attractive for the purposes of selective dissemination of information. SDI systems depending upon these sources of input tend to form natural clusters of similar configurations based upon the constraints imposed by the format of the input. Thus, systems depending upon CAS inputs tend to be similar, just as those depending on DDC, NASA or NLM inputs.

What still remains to be done is the establishment of meaning ful criteria for evaluating SDI systems operations. Different systems currently collect different operational statistics, so that valid conclusions cannot always be drawn from a comparison of published figures. As an example of the variability in reporting, some systems report relevancy as the ratio of the number of SDI notifications marked "relevant" by the recipient to the total number of notifications returned by the recipient. Other systems report relevancy as the ratio of the number of notices marked "relevan by the recipient to the total number of notices sent to the recipient. Thu two different systems reporting the same relevancy ratio might differ drastically in the actual relevancy of their respective system's outputs.

B. DATA INPUT

There are three types of input to an SDI system. These are:

* Information sources
* User profiles with addresses
* User responses

Information sources can consist of full texts of internal reports and memoranda, technical reports, books, journal articles, patents and symposia, o derivatives of these sources in hard copy or machine-readable form. The d rivatives may consist of titles, sources, corporate and personal authors, source numbers, abstracts and index terms. User profiles may be in the fo of a full-text description of information requirements or a description structured from index terms. User response may consist of index terms to used in modifying a profile, requests for documents, expression of the degree of relevance of notices received, change of address or general commen on the system's operation.

1. Information Sources

(a) Machine-readable document derivative data

There is a wide latitude in the types of information sources accepted by the various systems studied. Since the data in machin readable form available from DDC, NASA, NLM and CAS contain only derivativ of full text documents, the systems which utilize these inputs accept, by definition, derivative information. Early in 1962, Chemical Abstracts Ser

commenced an experiment to disseminate Chemical Titles on magnetic tape to Eli Lilly and Company and Olin Mathieson Chemical Corporation. The tapes contained only title, personal author, and journal source. Chemical Titles covers 590 important U. S. and foreign chemistry and chemical engineering journals. Since some of the source data for Chemical Titles is derived from journal page proofs, some of the titles appearing in Chemical Titles precede the actual publication of the articles represented. Thus the advantages of using Chemical Titles information are the breadth of coverage and the timeliness of announcement. A distinct disadvantage is the lack of control over vocabulary; user profiles must be phrased to include all possible synonyms which must be derived through cerebration, rather than table look-up in a published thesaurus. Because of the synonym problem, user profiles tend to be much larger than average, some extending to several hundred terms. To circumvent this difficulty, most systems using Chemical Titles have provided for the inclusion of truncated terms, as, for instance, separat for "separate," "separator," "separating," "separation," or "separated."

The proven utility of the tape distribution program prompted Chemical Abstracts Service to offer magnetic tapes to the general public on a fee basis. The program was expanded to include also Chemical-Biological Activities magnetic tapes. American Cyanamid Company and Dow Chemical Company are among several other organizations which have joined the program. Study is currently under way to determine the feasibility of including full abstracts from Chemical Abstracts Service tapes, but these tapes would not have the same advantage of timeliness as Chemical Titles and Chemical-Biological Activities.

Magnetic tapes from DDC and NASA contain all the information that is printed in the TAB and STAR announcement journals, respectively. NASA's SDI program began by servicing individuals and groups engaged in aerospace technology. SDI notices were sent to both NASA personnel and contractors. Recently, NASA has released magnetic tapes to several computer installations, one of which is located at the University of Pittsburgh. The University now offers an SDI service on a fee basis to government, university and industrial subscribers. NASA encourages the heaviest users of its SDI system to obtain NASA tapes in order to provide their own SDI service, internally.

On May 19, 1966, DDC held a seminar for DOD contractors and government agencies on the format and contents of its magnetic tapes. Although this conference was not held for the specific purpose of encouraging the use of DDC tapes for SDI, many of the participants expressed interest in this particular application. The U. S. Army Electronics Command at Ft. Monmouth, New Jersey, had already negotiated the release of DDC tapes which they planned to use to implement an internal SDI system. Programming was already underway to achieve that end.

The U. S. Army Biological Laboratory at Ft. Detrick, Maryland, obtains periodic magnetic tape output from the National Library of Medicine, and is presently using the tapes as a data base for SDI. These tapes are similar to CAS tapes in that they do not contain abstracts, but do

contain a list of subject headings by which each document is indexed in the MEDLARS program. The inclusion of subject headings facilitates sea ing and profile maintenance because the MEDLARS subject heading list ca be used for selecting unique terms for profiling without necessitating addition of all possible synonyms.

Thus far, SDI systems utilizing externally produced machine-readable data rely on a sole source, even though several source might supply equally useful machine-readable input. A logical reason f this is that these systems are in an early stage of development and are now completing the debugging phase of systems development. Once a sing machine-readable input has been mastered, assuredly other machine-reada inputs will be added.

(b) Non-machine-readable input

Those systems which rely on externally-produced mag tape as input are automatically committed to disseminating information many sources, since these tapes already contain derivative data from jo al articles, books, technical reports, patents, etc. For those systems which depend on original input, the choice is still open. Some systems as Douglas Aircraft Company, Mitre Corporation, and two systems formerl erated by IBM (Owego and Yorktown Heights), have elected to limit input solely to documents which stand by themselves in the form of contractua documents, or journal and symposia preprints and reprints. Full journa proceedings of symposia are not included because of the difficulty of i ing individual articles for distribution. Internally-generated reports disseminated by the present system at IBM (Yorktown Heights), National nautics and Space Administration, U. S. Air Force Systems Command, and Department of the Interior's Bureau of Reclamation. Both the Institute Scientific Information and IBM (Yorktown Heights) disseminate patents. the other hand, many systems rely totally or in part on the input from journal articles. These include Chemicals and Phosphates, Ltd. (Israel IBM (Yorktown Heights), Institute for Scientific Information, U. S. Air Force Systems Command, and, of course, NASA, DDC, NLM and CAS.

As pointed out above in the discussion of the basis design of an SDI system, the choice of inputs is governed by the inform requirements of the users, the adequacy of coverage by other informatic services, the potential volume of input, and the resources available fo operating the SDI system. In some instances, the selection of input ha so been governed by copyright laws which protect much of the open liter such as books and journals. Since the reproduction and distribution of lished abstracts may constitute a copyright infringement, some systems refrained from wholesale input of abstracts readily available in publis abstract and announcement journals. A recent journal article (see Appe C - Reference No. 7) discusses the presently proposed changes in copyri laws which are necessitated by the high usage of copyrighted material i formation retrieval systems. Some of those systems which copy publishe abstracts have made special arrangements with the publisher to avoid co right infringement suits. This problem has also arisen in conjunction

the reproduction of published author abstracts by DDC and NASA. Even though these agencies may have obtained permission to reprint author abstracts, the secondary reproduction of these abstracts by various SDI systems still may constitute infringement. Although reproduction of abstracts might constitute fair use of copyrighted materials, there is no doubt that reproduction of full journal articles and other copyrighted documentation is unfair use and that publishers generally object to the wholesale copying practices of many public and private libraries.

(c) Preparation of original machine-readable input

Those systems which do not accept externally generated magnetic tapes as input must operate their own data input system in which part, or all, of the document derivative data is punched into cards or paper tapes. Organizations such as the Bureau of Reclamation, Douglas Aircraft Company, and the U. S. Air Force Systems Command use punched paper tape for input to the computer and for direct creation of other library records such as catalog cards and accessions lists. All other systems studied enter document derivative data on punched cards and use a computer-driven, high-speed printer to generate user notifications only, or both notifications and library records. Using a 400 line-per-minute printer to print 20-line user notifications by means of a computer renting for $120 an hour, the printing cost per notice is 10¢, not including the cost of forms or the cost of putting the input information in machine-readable form, or processing time for matching document descriptions against user profiles. It can be seen from this simple calculation that using the line printer for printing multiple copies of each set of document derivative data is rather expensive compared with the cost of printing by conventional office duplication or offset press techniques. Preparation of punched card input is also approximately three to four times as expensive as the preparation of typewritten offset masters. The ratio is even higher if document derivative data is available in a format which can be extracted by photoduplication techniques. In striving to achieve a totally automated SDI system, most designers have overlooked the break-even point in economics between printing by computer-driven line printer and printing by offset techniques.

(d) Subject indexing and abstracting of input

Of all systems studied, only IBM performs automatic indexing on full abstracts. Those organizations using CAS tapes must, of necessity, auto-index the titles in order to derive terms for matching against user profiles. All other systems either utilize the subject indexing performed by others, or do original subject indexing. If externally-assigned index terms are accepted into the system, then profiles must be drawn from the thesaurus from which these index terms were assigned. The U. S. Army Biological Laboratory uses Medical Subject Headings (MESH), published by the National Library of Medicine, to be compatible with the subject index terms assigned to document derivative data obtained from the MEDLARS program. The U. S. Army Electronics Command and the system formerly operated by IBM (Owego) used the DDC thesaurus. Chemicals and Phosphates, Ltd., the system formerly operated by IBM (Poughkeepsie), and Dow Chemical Company utiliz

the thesaurus published by the American Institute for Chemical Engineering. Some systems, such as the Bureau of Reclamation, NASA, the U. S. Air Force Systems Command and Douglas Aircraft Company, have created their own unique thesaurus, tailor-made to meet the subject requirements of the users of their entire range of information services. Although many of these unic thesauri have been built by extracting parts of other published thesauri, the extracted terms take on new significance through usage and thus cannot be used to advantage in retrieving materials indexed by the original thesaurus publisher. For example, DDC uses the term "fibers" to describe documents on both single crystal whiskers, as in the title "Reinforcing Effects of Silicon-Nitride Whiskers in Silver and Resin Matrices," as well as documents on spun fibers as in the title "Fiber Formation in Web Spinning c Polyacrilonitrile."

Some systems which utilize the CAS tapes will allow selection to be based on the presence of author's name and source journal. The service commercially offered by the Institute for Scientific Informatic is based entirely on the ability to select specific reference cited, specific author cited, specific author published, author affiliation and patent class and sub-class. There is no facility for searching journal articles t subject description. There is no reason why contract numbers, report serie acronyms and other such descriptive cataloging information could not also t utilized in both document and user profiles.

The system formerly operated by IBM (Chicago) allowed the system's operating personnel to route documents to certain individuals by job classification. For instance, the profiles of various salesmen were tagged with the term "salesman", and any document which was thought to be c interest to these people was similarly tagged. When the document derivati data was matched against all user profiles, all documents indexed by the term "salesman" would automatically be distributed to salesmen on the basis of the single matching term. In Luhn's original system description, he mad provision for the user himself to suggest other users who might be interest in receiving the same notification which he received. The capability for referral of this type was not implemented by any other system examined thus far. No other system besides IBM (Chicago) and Luhn's original system take advantage of the potential for introducing the human element in the SDI routing process.

In addition to performing original indexing, some systems also assign weights or links and roles to index terms. The Bureau of Reclamation imitates DDC's practice of attaching an asterisk to those index terms which describe the document contents most aptly. Chemicals and Phosphates, Ltd. assigns a role code to each descriptor. Thus, water can be coded as a solvent or a medium. An experimental modified link system was used by one of the Douglas Division libraries, but the time spent in accurately linking descriptors was a more significant problem than the false drops occurring during the several years of program operation. For this reason, the use of links by Douglas was discontinued. No system has used a numerical weight assigned to each descriptor to describe the degree to

which the descriptor is applicable to the particular document, but the Bureau of Reclamation and the U. S. Army Materials Command use an asterisk to indicate the most pertinent descriptors.

The above discussion can be summarized in the following statements which emphasize the variability of the systems studied.

(1) Few systems accept externally-produced magnetic tapes.
Most systems prepare their own machine-readable input.
Most systems use punched cards.

(2) Most systems disseminate only externally-authored information.
Few systems disseminate internally-produced information.

(3) Most systems put all document derivative data, including an abstract, into machine-readable form.

(4) Most systems index documents by means of terms selected from a carefully controlled thesaurus.
Few systems rely on auto-indexing.

(5) Most systems accept information from multiple sources such as journals, books, technical reports, patents, etc.
Few systems input only one source.

(6) Most systems input information which is normally scanned or readily available to the system's users.
Few systems cover journals or other sources which contain a relatively low percentage of pertinent information.

(7) Most systems do not exclude input information on the basis of original publication date, but most systems process a high percentage of recently published material.

Since many of the choices exemplified by the above differences in systems input have been made on the basis of personal preference or expediency, there is no reason why the majority approach should be adopted in designing a new SDI system.

2. User Profiles: Construction and Effect on Document/User Profile Matching

User profiles are predominantly drawn by the combined effort of the user and personnel associated with operation of the system. Long-term subscribers who have gained experience in the relationship between their profile and document profiles are capable of drawing satisfactory profiles without the aid of system's personnel. New subscribers, however, must be guided very carefully in the selection of terms to be used in their profiles; otherwise the results of dissemination tend to be narrower or broader than desired. Most systems have a prepared form on which the user can submit a profile in either descriptive style or in the form of specific index terms.

Of necessity, the structure and composition of the user profile must parallel the structure and composition of document profiles prepared on the information input side of the SDI system. In preparing profiles to be run against CAS magnetic tapes, the user must supply all possible synonyms for the word or words that describe the concept in which he is interested. To keep profiles to manageable size, CAS developed a computer program that would search for the occurrence of any title word beginning with a certain string of letters. This eliminated the necessity for specifying the same profile word with all possible endings. Former IBM programs also allowed truncation of terms.

Due to the influence of Taube's original Uniterm concept, several systems originally limited index terms to one word, such as the early Douglas system, but most systems now accept multiple-word terms in user profiles and use multiple word terms to index documents. In the current IBM system and the Chemical Abstracts system, the user can stipulate that one term must follow another in order for a match to occur. In the Chemicals and Phosphates, Ltd. system, the user can specify a certain role for each term in his profile. In those systems using weights, each profile term must be associated with a given weight value. In earlier IBM systems, weights ranged from -9 to +9. The Iowa State University system uses a weight ranging from 0 to 1. The Bureau of Reclamation system allows the user to specify four descriptors with an asterisk to indicate the highest degree of importance. Those systems utilizing weights include the Douglas Aircraft Company, Dow Chemical Company, former IBM (Chicago) system, Iowa State University, U. S. Air Force Systems Command, and the Bureau of Reclamation.

In systems using a Boolean search strategy, the Boolean combination of terms must also be specified in the profile. Boolean search systems include those of Eli Lilly & Co., the former IBM (Poughkeepsie and Yorktown Heights) systems, the NASA system and the present IBM (Yorktown Heights) system. Although weights are used in the U. S. Army Biological Laboratory system, the weights are assigned and profiles constructed in a manner which implements a Boolean search.

In those systems using a sum of weighted terms, the profile must also contain a numerical value to indicate the threshold which will cause selection, if exceeded by the sum of the weights of matching user profile and document descriptors. In all systems but one which use numerical weighting, the weights of matching user profiles and document descriptors are summed algebraically and compared against the threshold value specified by the profile. If the sum of the weights exceeds the threshold value, that document is selected for dissemination to that user. The one exception is the Iowa State University system in which the descriptor weights range from 0 to 1 and are interpreted as the probability that a document containing one of the weighted descriptors will be pertinent to the individual whose profile is being matched. In this system, two descriptor weights are combined by subtracting their product from their sum. In this way, the weight for each successive matching descriptor is added to the matching descriptor weights combined previously. This formula computes the joint probability of independent events. Current research in lexicography and information retrieval, however, points out that the occurrence of one descriptor may not be independent of the occurrence of another descriptor. For instance, if the term "radar" occurs in text, the probability that other words like "set," "antenna," "range," is higher than in average text.

The Bureau of Reclamation and the U. S. Army Electronics Command are the only systems using a two-valued weight for both profile and document descriptors. In the Bureau of Reclamation system, the user is permitted to assign a high weight to four of a maximum of twenty profile descriptors. In experimental trials of the Army Electronics Command system, every profile descriptor is assumed to carry a high weight. In both systems, documents are indexed with both high and low weight descriptors.

In Luhn's original system concept, selection was made on the basis of a percentage of document descriptors which matched the user profile descriptors. In this system, the user specified in his profile the percentage factor which would determine selection. IBM (Owego and Poughkeepsie) followed this match criterion. Due to the marginal success of this method, however, systems presently in operation do not use a percentage of document descriptors matching profile descriptors as a selection criterion.

In systems using a Boolean match function, the user must specify which terms are to be "OR'd" and which terms are to be "AND'd" together. These systems include the present IBM system at Yorktown Heights and the systems at Eli Lilly and Co., NASA and the U. S. Army Biological Laboratory, although the latter selection is implemented by descriptor weights.

In both the present IBM system at Yorktown Heights and the former systems at IBM (Owego and Poughkeepsie), the user can specify terms which must occur in the document description or terms which should not occur, if selection is to take place. Thus, no matter what the sum of the weights or the percentage of match, a document will be selected if there is a match between document and profile descriptors which have been marked as "must" terms and a document will never be selected if there is a match between descriptors marked "not".

In all systems studied, user profiles in each system have tended to cluster to varying degrees. In some systems, the clustering is so clearly defined that several user profiles could conceivably be combin into one profile. Both the Mitre system and the U. S. Air Force Systems Command utilize this phenomenon by writing profiles for groups as well as for individuals. This saves computer time because it reduces the total n ber of profiles that must be searched. In many instances, the similarity between profiles can be ascertained by human inspection, but in systems servicing thousands of users, such as that operated by IBM, manual inspec tion would be too cumbersome. Statistical procedures have been developed for pattern recognition that might be useful, also, in determining the de gree of similarity between profiles. This might be a profitable area for further study because of the potential reduction that might be obtained i the total number of profiles contained in a single system.

In many instances, the similarity between profiles can be a certained by visual inspection. But in systems servicing thousands of us such as that operated by IBM, manual inspection would be too cumbersome. Statistical procedures have been developed for pattern recognition which might be useful also in determining the degree of similarity between profiles. This also might be a profitable area for further study.

3. User Response

By definition, an SDI system must allow for adjustment of u profiles on the basis of the response received on previous SDI notices. degree of automation of this SDI function ranges from the completely auto mated profile updating accomplished by the Iowa State University system t the completely manual system operated in a pilot test by the U. S. Army Electronics Command. The Mitre Corporation system represents a unique ca because provision for modification of user profiles was built into the sy tem but profiles are not currently being modified.

Most systems provide the user with a printed form for indic ing the relevance of each SDI notice sent to him, and require that the fo be returned to the system whether or not the notice was relevant. In the Bureau of Reclamation system, however, the user is required to return the notice only if it was pertinent. Most of the forms contain at least four positions for indicating whether the notice was

(a) of interest, document requested,

(b) of interest, document not wanted,

(c) of interest, have seen before, or

(d) of no interest.

The forms used by Iowa State University, IBM (Yorktown Heights, Chicago, and Owego), and the form formerly used by NASA, were constructed in such

way that the user could punch out a hole next to the appropriate comment. The manually punched EAM card (Port-A-Punch card)* could then be processed mechanically by EAM equipment.

In most systems the response form and the SDI notice are printed in one operation on a high-speed line printer, but the exceptions to this are the Bureau of Reclamation, which prints both notice and response cards by offset techniques, and the U. S. Army Electronics Command, which prints by standard office duplicating techniques. In those systems which provide the user with a manually punchable response form, the SDI notice and form are printed on two separate IBM cards which fit side by side on continuous-form-feed card stock. The Bureau of Reclamation system prints the response form and SDI notice on one IBM card, which is perforated to facilitate separation of the notice from the response form. In the Bureau of Reclamation, the former NASA, USAF Systems Command, Douglas, and U. S. Army Biological Laboratories systems, the user notification is printed in a 3" x 5" card format which can be separated along pre-perforated lines from the response form.

Only Iowa State University uses the user response form to adjust automatically the weights assigned to user profile descriptors. All other systems require the user to indicate addition, deletion, or modification of profile terms by writing these changes on the response form provided, or by requesting the changes in a letter format. In all systems except Mitre Corporation, the user can request changes in his profile by specifying the addition or deletion of terms, by changing the weights assigned to terms, or by specifying new threshold values. Systems using a Boolean search strategy allow the user to modify the Boolean combination of descriptors in his profile.

In all systems, provision has been made for maintaining an up-to-date address list. Lack of feedback of change of address information is a problem in every system. In most systems, the user address is maintained in machine-readable form with each user profile so that addressing of user notices can be accomplished simultaneously with printout of each user notice.

Users respond with general comments as well as specific profile modifications and address changes. These occasion a personal consultation with the user, if the response is very negative. In the U. S. Army Biological Laboratories and IBM (Yorktown Heights) systems and other systems of like magnitude, users are periodically contacted to solicit comments on the utility of the system. These comments may be used in implementing modifications indicated by a majority of the users. All systems tabulate some statistics which are used to single out individuals who may be experiencing some difficulty with the system.

* IBM trademark

C. DOCUMENT/USER PROFILE MATCH STRATEGY

There are three major search strategies used by SDI systems. These can be described as: (a) Boolean, (b) MUST, NOT, MAY, and (c) weighted. Some systems use a combination of two of these search strategies.

A Boolean search strategy is one which combines user profile terms in a Boolean function whose logical operators are AND, OR and NOT. The AND operator will cause selection if, and only if, both terms in the expression are present in the list of document descriptors. The OR operator will cause selection if one or the other or both terms in the function are present in the document descriptors. The NOT operator will prevent selection if the negated term is present. In all systems using the Boolean search strategy, user profiles are constructed from lists of descriptors which are AND'd together or negated. Within each list the descriptors are OR'd together. Thus, a question takes on the form (a or b or c) and (d or e or f or g) and (not h or not i), where a, b, c, etc. are descriptors in the profile. Chemicals and Phosphates, Ltd. and Mitre Corporation permit only one list of terms per user profile so that a document is selected if any one of the terms in the list is present in the document descriptors. Eli Lilly & Co., NASA, IBM (Yorktown Heights) and U. S. Army Biological Laboratories permit the construction of more complex Boolean functions. The Institute for Scientific Information also uses a single-list Boolean search strategy, but the terms OR'd together consist of personal author names, journal sources and corporate sources rather than terms selected from a controlled vocabulary.

The earliest system developed by Luhn and the IBM system formerly operated at Poughkeepsie used the MUST, NOT, MAY strategy in which selection was based on a percentage of document descriptors which matched the user profile descriptors. The user could also specify an absolute selection or rejection on the basis of the presence of certain descriptors. Both MUST and NOT terms override MAY terms. NOT terms override MUST terms. This strategy achieved marginal results because of the apparent unpredictability of the percentage-of-matching-terms selection criterion. Also, the ability to select or reject on the presence of a single descriptor put undue importance on that descriptor. The user had to balance the undesirability of obtaining too much information against the desirability of obtaining everything which was marginally pertinent. Many documents were rejected because of the presence of negated terms which would otherwise have been selected because of a high percentage of matching descriptors or the presence of MUST terms.

By far, the most popular search strategy involves the assignment of weights to each profile descriptor. In the systems operated by Douglas Aircraft Company, Dow Chemical Company, the former IBM (Chicago), the U. S. Air Force Systems Command and the U. S. Army Biological Laboratories, the weights assigned to those user profile terms which matched document terms are summed algebraically and compared to a threshold value. If the sum of the weights equaled or exceeded the value of the threshold, the document was selected. If the sum of the weights was less than the threshold value, the document was rejected.

Of the three search strategies, sufficient evidence is present to conclude that the MUST, NOT, MAY search strategy produces a system which is ultrasensitive to the balance between relevancy and recall. By very slight changes in the profile, a high degree of relevance can be obtained with a high loss of recall. In order to improve recall, one must sacrifice relevancy to some degree. With the Boolean search strategy, one can obtain a higher degree of relevancy with better control over recall at the same time. The use of weights yields a high recall with little sacrifice in relevancy.

The various IBM systems introduce random SDI notices in order to test the system's efficiency. Random outputs are generated under the theory that a high percentage of positive responses to random notices indicates that the particular user profile is incurring a low recall; that is, the percentage of relevant information retrieved from the entire collection is too low. There may also be some merit in introducing random notices to expose the user to a cross section of all documents being put into the SDI system so that the recipient can modify his profile to reflect a broader spectrum of information available in the system.

If a system of weights is used in the selection strategy, the sum of the weights could be used as a relevance score by which selected documents could be sorted from highest-scoring selection to lowest-scoring selection. Then, the top ten or so selections could be printed out for notification. This would be a feasible method for limiting the total number of selections per user for each SDI cycle. Studies reported by A. Resnick and C. B. Hensley (see Appendix C - Reference No. 67) and A. Resnick (see Appendix C - Reference No. 66) indicated that approximately two-thirds of the subjects tested did not want to place a limit on the number of SDI notices received per week. The mean number of notifications received per week per user, however, was 3.5, although there was a large variance among users. It is probable that, in systems processing a large number of documents per SDI cycle, some of the users would want to limit the number of notices sent to them.

The present IBM system at the Thomas J. Watson Research Center, Yorktown Heights, N. Y., and some of the systems based on input from Chemical Abstract Service, allow the user to specify word roots and word sequences as part of the Boolean function comprising the user profile. This facility is useful in matching profile terms selected from an uncontrolled vocabulary against terms found in normal text. By allowing the user to specify that word B is to follow word A, document selections are prevented in which word A follows word B. As discussed under user profile input, word roots are useful for cutting down the size of the user profile, because a single character sequence can be used to express variations that occur in normal text. The ability to detect word sequences and word roots is only necessary to those systems such as IBM and CAS which search full text abstracts or unedited titles.

Relevance and recall are two measures of systems efficiency which have often ben applied in the analysis of retrospective information retrieval systems. These measures are equally applicable to the analysis of SDI systems. Relevance is defined as the ratio of the number of relevant documents retrieved to the total number of documents retrieved. Recall is defined as the ratio of the number of relevant documents retrieved to the total number of relevant documents in the system. The relationship between relevance and recall is such that any attempts to improve relevance usually produces a loss in recall. To improve relevance, one must be more specific, but this can only be accomplished by narrowing the scope of the search. On the other hand, if one wants to be assured that he is obtaining all relevant documents in the collection, he must specify a broader search, so that documents will be selected that might be on the fringe of his area of interest. By experience, it was found that the MUST, NOT, MAY search strategy was ultrasensitive to the balance between relevance and recall. For this reason, the MUST, NOT, MAY strategy is no longer popular as a sole search strategy, although it is still used in some systems in conjunction with the weighting strategy or the Boolean strategy. The Boolean strategy has the capacity for obtaining high relevancy with tolerably good control over recall. The weighting strategy is best suited to producing high recall with fair control over relevancy. On the basis of relevancy figures published by the various systems, one cannot conclude that the Boolean strategy is significantly superior to the weighting strategy or that the weighting strategy is significantly superior to the Boolean strategy because the number of variables in each system might account for local success with one system or the other. The user preference for the results obtained from using one system or the other is very much dependent upon the activities engaged in by the user. As an example, administrative personnel might be more satisfied with a high degree of recall and a reasonable level of relevance whereas research personnel might demand high relevance with a corresponding sacrifice in recall.

D. DATA OUTPUT

Data output consists of notices sent to the user, hard copies supplied on demand, and statistical reports. Among the systems studied, there is an extreme degree of variability in these three kinds of outputs in terms of both quantity and content. Each of these outputs is discussed under separate headings below.

1. User Notifications

All systems included, as a minimum, the title and descriptive cataloging information for each document. The latter includes the author, corporate or journal source, and report number. In systems based on CAS or NLM inputs, this is the only information provided on the user notice since only this information is present on the input magnetic tapes. Most systems, however, also provide an informative or indicative abstract. An indicative abstract tells the reader what elements of information are contained within the document, whereas an informative abstract provides some

of the information itself. The Bureau of Reclamation's system is most noteworthy for providing informative abstracts of considerable length, i.e., averaging well over 200 words per abstract. Other systems provide informative or indicative abstracts as they occur in the documents themselves.

For the most part, where abstracts are not provided with the full text document, most systems create an indicative abstract. IBM uses trained secretaries to create abstracts, whereas the Bureau of Reclamation parcels out the abstracting task to members of the engineering staff which is served by the SDI system.

Most systems also provide either the full set of descriptors by which the document is indexed or the set of descriptors which cause selection. The former IBM (Chicago) system and the Iowa State University system provide the latter. All other systems give the complete descriptor list. The inclusion of descriptors on the user notice provides the recipient with an indication of the reason for selection of that particular document. If the document notice is irrelevant, the user can request that those descriptors which caused selection be eliminated from his profile. The system formerly operated by NASA did not include index terms, but these were later incorporated in the notice printed by their new system. The University of Iowa system also gives the weights assigned to all of the matching descriptors.

The choice of what informative information to provide on the user notice must be guided by the objectives which the system is to accomplish. When input is accepted from outside sources, no more information can be provided on the notice than is present on the original input. But for those systems which create original input, the system designer has the option of disseminating outputs ranging from self-sufficient information packages to mere indications of the existence of those packages. The Bureau of Reclamation has chosen the former route, whereas most other systems have chosen the latter, even though they provide abstracts.

As mentioned previously, all systems except the U. S. Army Electronics Command and the Bureau of Reclamation print the user notification and response forms directly on a computer driven line printer. Both the Institute for Scientific Information and NASA print on a paper form, whereas all other systems print on card stock forms. The most popular card stock form size is the standard EAM card. The utility of the notice to the user is much enhanced by providing a notice that can be reduced to 3" x 5" card size for manual filing. The Bureau of Reclamation and the U. S. Army Biological Laboratory provide notices in this form. Formerly, NASA and IBM printed in a 3" x 5" format, but discontinued this practice for various reasons.

The majority of systems print a one- or two-line address directly on each notice. Since the notices are printed in user profile number order, the collation and packaging of all notices going to one user is accomplished merely by bursting the continuous form feed output and separating the stack of cards or forms into groups by profile number. Since

notices are printed in document number order by the Bureau of Reclamation and the U. S. Army Electronics Command, these must be sorted by hand into profile number order in order to collate and package the notices going to one user.

2. Statistical Reporting

The most commonly reported statistical information consists of tabulating the number of user responses of each type. Most systems solicit four types of responses (i.e., of interest, document not wanted; of interest, have seen before; of interest, document requested; and of no interest). A refinement of this type of reporting consists of adding together the number of responses in the "of interest" categories and dividing this figure by the total number of notices sent to obtain an average relevance percentage. Most systems also compute a relevance percentage for each user, so that unusually low relevance rates can be detected and corrective measures taken.

For the systems listed below, the following statistics are reported in the documentation listed in Appendix B.

Bureau of Reclamation

- Total notices sent
- Total responses received
- Total number of documents ordered

Dow Chemical Company

- Number of notices of interest
- Number of notices of some interest
- Number of notices of no interest
- Number of comments

IBM - Present System

- Total number of notices sent
- Number and percent of responses returned
- Number and percent of interest, with breakdown into three categories
- Number and percent not relevant

Iowa State University

- Percentage of descriptors contributing to document selection
- Percentage of notices of interest
- Percentage of documents ordered
- Number of input documents
- Number of documents selected for dissemination

National Aeronautics and Space Administration

Percentage of document notices marked relevant

U. S. Air Force Systems Command

Number of selections per profile term

U. S. Army Electronics Command

Average number of notices sent to each user
Average number of notices of interest
Average number of documents ordered

Almost all of the systems studied keep statistical records which reflect the operation of the system, but these were certainly not freely reported in the open literature, except for a very few of the systems. Relevance figures reported range from approximately 50 to 70 percent. Although the former system at IBM (Poughkeepsie) reported a consistent 75 to 80 percent relevancy of notices sent, this figure seems unusually high and might actually indicate the percentage of notices returned which were marked relevant. This interpretation would lead to a relevancy figure consistent with that reported by the present IBM system at Yorktown Heights. The Yorktown Heights system reports that 81.1 percent of the notices returned were marked relevant for the notices sent describing IBM documents. Relevancy computed on the basis of the total number of notices sent on IBM documents yields a percentage factor of 45.2. As mentioned previously, the two methods of computing relevancy yield two extremely different results, so that two systems cannot be compared on the basis of relevancy, if two different computation methods have been used.

The present IBM system automatically detects users who have received no notifications in the current run, users who failed to return their response cards within a specified period, and users whose nonrelevant response exceeded a predetermined percentage. Other systems have similar features to detect conditions which might indicate that modifications should be made in the user's profile. Iowa State University and the U. S. Air Force Systems Command detect the number of profile terms that have contributed to the selection of notices. This information is useful in determining which descriptors infrequently contribute to selection so that these can be weeded out from profiles. Several systems periodically issue statistical reports to the user to indicate how the user's response and use of the system compares with average uses and responses.

3. Copy Supply

The document copy supply problem is ever present in SDI systems, since many requests for the same document are received simultaneously shortly after the conclusion of each SDI cycle. The problem is met in different ways by different systems. IBM puts all input documents on microfilm

and distributes a full microfilm collection to each location served by the SDI system. Microfilm copies offer the least expensive reproduction method, if copies must be distributed to many locations. Once the master microfilm record has been created, subsequent microfilm copies can be reproduced at a price in the vicinity of twenty cents per document.

Most systems rely on standard office duplicating techniques for reproducing small numbers of copies of documents ordered. Price per page on a Xerox 914 machine is five cents, not including labor. In systems disseminating information on documents not on hand, orders for documents must be fulfilled on an individual basis by ordering from the source. For those contractors or Government agencies utilizing Defense Documentation Center or National Aeronautics and Space Administration, document copies can be obtained at no charge, but a delay is incurred in ordering.

For those systems which disseminate information on journal articles, those journals which demonstrate a high popularity can be ordered in multiple copies. Chemicals and Phosphates, Ltd. of Haifa, Israel appends a circulation list prepared by the dissemination system to each journal issue. The obvious disadvantage to this system is that the last person on the list must wait a considerable length of time before viewing the journal article. If copyright laws are strictly observed, journal articles should not be copied without permission from the publisher. Some systems ignore the document supply problem by stating at the outset that the user is responsible for obtaining his own document copies from whatever source he chooses and by whatever methods he has at his own disposal.

The number of documents ordered can be reduced by supplying more information on the original dissemination cycle. The U. S. Air Force Systems Command will provide, at the user's option, document derivative data only, document title and abstract only, full document only, or any combination of these three.

E. COST ANALYSIS

The only system for which a complete cost analysis could be obtained was the IBM SDI-3 system written for the IBM 1401 digital computer. Estimated monthly costs were given to process 50 abstracts per day.

Fixed costs were:

Personnel	$4100.
3 IBM 826 typewriter card punches	480.
Punched cards	50.
Contingencies	500.
Total	$5130.

Variable costs per 100 profiles at 5 notices per person per day were:

8K 4 Tape IBM 1401 (5 hours per day on extra shift, 40% of prime shift rental)	$2200.
Forms	1100.
Photographic copies (300 sheets per day)	800.
Contingencies	300.
Total	$4400.

Based on these costs and conditions, the estimated cost per profile per month is $9.53. As the number of profiles goes up, the fixed costs quoted above are distributed over the larger number of profiles to produce a lower cost per profile. Thus, for 2000 profiles, the unit cost is $7; and for 3000, the unit cost is $6.10. The cost per year for processing 12,000 documents for 500 profiles amounts to $92,40.

A reflection on these high costs is evidenced only by cryptic remarks in the literature. Quoting from a paper published by American Cyanimid (see Appendix C - Reference No. 26) describing the use of SDI-3:

> "The preparation of input presented some difficulties. One peculiarity of the program imposed certain restrictions on abstracting procedures. Profiling terms of two or more words must be coded to be read as one word. Unfortunately consecutive word terms in an abstract also require the same special coding in order to match this type of profile term. As a result use of chemical names of profile terms must be very carefully handled. The clerical work involved in keypunching and verification also added to the input problems. The costs involved have been very high. For these reasons such a system could hardly be considered a solution to a small information center problem."

Quoting from another source (see Appendix C - Reference No. 24) which describes the use of CAS tapes by Olin Mathieson Chemical Corporation:

> "Although we do not yet have enough data to draw final conclusions, the most favorable combinations of the four parameters (which contribute to costs) do not appear to justify the present cost of machine search. However,

considerations of speed, available staff, and the alternative uses for the time consumed, if each scientist were required to make his own searches, may justify the cost for machine matching. The cost of operating second generation systems we contemplate may be less than that of a manual system."

The above discussion should be sufficient to indicate that cost not a trivial consideration in establishing an SDI system. Unit costs profile are considerably reduced if the fixed cost of input preparation distributed over a large number of profiles. For a system servicing a number of users, an attempt should be made to reduce fixed costs, since these would contribute most to the cost of servicing each profile. On other hand, a system which services a large number of users should stri to reduce the variable costs, which constitute a fixed cost per profile

Fixed costs are composed of the following elements: personnel pervision, abstracting and indexing personnel, clerical personnel, keyp or paper tape typewriter operators, materials, and contingencies. Maxi utilization of pre-abstracted documents would contribute substantially reduction in fixed costs. If a reproducible master is prepared by phot copying techniques from already published abstracts, this would contrit an additional cost reduction. Of course, not supplying the abstract at would entirely eliminate the cost of creating and/or transcribing an ab stract. But this route might be undesirable in terms of overall systen objectives.

Variable costs are composed of these elements: computer proce: ing time, materials, reproduction, and contingencies. Computer time i: sumed in record-keeping functions, in execution of the match function, in producing various outputs. Of the three, the time consumed in actu: selecting the information to be disseminated is neglible in comparison the other two. Without sacrificing systems objectives, the only area i which considerable savings could be achieved is in the area of output p ing. The following comparative table shows the cost of operating the p ing cycle for 3000 input documents disseminated to 600 users who recei average of 150 notices (containing an average of 1000 characters, each tributed over 20 lines of text):

Computer Printing Costs	Offset Printing Costs
Key and Verify	
- on keypunch	- on typewriter or Flexowriter
3000 abstracts @ 95¢=$2850.	3000 abstracts @ 35¢=$1050.
Record	
- on magnetic tape	- on offset masters
5 min. @ $120 /hr. =$ 10.	3000 masters @ 10¢ =$ 300.

Print		
- on 1000 line per minute on-line printer	**- on offset press**	
30 hrs. @ $75./hr. =	$2250.	15 hrs. @ $30./hr. = $ 450.
TOTAL COST	$5110.	$1800.
Cost Per Thousands of Notices	$56.78	$20.00

Since most systems print the notices in users' sequence, no further work must be performed in collating the order for each user. When cards are printed by offset press, however, as in the Bureau of Reclamation system, the cards come off the offset press in document number sequence and must be collated into orders for each user. This would be an additional cost to the system which prints by offset.

The following times are quoted by the former IBM Chicago system on the IBM 1401 computer:

16 seconds per abstract to auto-index, using an exclusion list of 600 common words

60 seconds per abstract to auto-index, using a dictionary of 2200 words and also an exclusion list of 600 words

.127 seconds to compare one average profile with one average abstract

1.8 seconds to print one average abstract

Douglas Aircraft quotes 15 minutes of IBM 7094 time and printing and assembling time of 1 hour for an SDI cycle conducted for a typical library division at two week intervals. This does not include the 45 minute run on the IBM 7094 which produces the magnetic tape input for the SDI program. This same program also produces output tapes for input to other programs which, in turn, produce the acquisitions list, catalog cards, and a list of typing errors. Eli Lilly & Co. quotes 4 hours of IBM 1401 time and 1.4 hours of IBM 705 time to process one July, 1964 issue of Chemical Titles for 43 individuals with 1.25 requests per user and a total of 2100 key words. Each issue of Chemical Titles lists approximately 4000 titles with author and journal sources.

Cost figures for most systems are non-existent. In some instances where computer operating times are listed, one can obtain some idea of total costs, but the cost of performing indexing and abstracting is often hidden in the cost of operating other library services. If computer operating times are taken from several systems, one obtains a better picture of the overall cost of operations. Such a comparison leads to the conclusion that the cost data quoted by IBM is on the low side.

In weighing the relative merits of various methods for accomplishing the individual functions of an SDI system, one must consider the number of users, the document input rate, and the average number of notices sent per user. These factors will determine the cost incurred in terms of inconvenience, if some of the operations are conducted manually. The inconvenience must then be weighed against the cost of performing the same operations by computer.

F. MEASURES OF SDI SYSTEMS EFFICIENCY

Three measures have evolved for the evaluation of SDI systems efficiency. These are the relevance and recall ratios, the dissemination ratio and the selection-participation ratio.

1. Relevance and Recall Ratios

These ratios are used often in evaluating the efficiency of the conventional retrieval system. The relevance ratio is the ratio of the number of references deemed pertinent to the total number of references selected. For instance, if the system retrieves 100 documents, but only 25 of them are pertinent, the relevance ratio is .25. The recall ratio is the ratio of the number of pertinent documents selected to the total number of pertinent documents in the system. For instance, if 100 pertinent documents are known to be in the system, but only 5 are retrieved, the recall ratio is .05.

The relevance ratio is fairly easy to compute, since the total number of documents retrieved is an obvious fact, and the number of relevant documents retrieved is purely a subjective measure. The recall ratio, however, is difficult to compute since the total number of pertinent documents in the system must be determined by either examining all of the documents or by appropriate sampling techniques.

2. Dissemination Ratio

The dissemination ratio is the ratio of the average number of notices sent to the total number of documents scanned. For instance, if 1000 documents are scanned and an average of 100 notices are sent to each participant, then the dissemination ratio is 10%. This measure indicates the screening capacity of the system. That is, 90% of the documents are eliminated from consideration by the user because of the operation of the system. This measure does not take into account the fact that none or all of the notices sent might be pertinent.

3. Selection-Participation Ratio

The selection-participation ratio is the ratio of the number of descriptors in the profile that participate in the selection of documents. For instance, if each profile contains an average number of 30 terms and only an average of 6 terms participate in selecting pertinent documents then the ratio is .2. A high ratio indicates that the profile is very

efficient with respect to the particular data base. An efficient profile indicates that each matching operation has a high probability of yielding output. If the ratio is small, it means that the computer is searching for a match against subject areas that, in all probability, are not represented in the data base.

In the ideal system, the following characteristics are desirable: relevance ratio high, recall ratio high, dissemination ratio intermediate, selection-participation ratio high. If the dissemination ratio were high, almost all of the input documents would be disseminated to each user. This would indicate that there was no need for a dissemination system, since the user could do almost as well by accepting the total input.

G. EFFECT OF SDI ON THE USER

The following discussion is based on "The Use of Diary and Interview Techniques in Evaluating a System for Disseminating Technical Information," by A. Resnick and C. B. Hensley, published in 1962. Although the paper states that "this study, though more sophisticated and better controlled than some reported in the literature, still falls short of the ideal of a carefully controlled scientific experiment," it still remains one of the best studies on the effect of SDI systems on the user. The experiment was conducted on 45 IBM research personnel who had no previous familiarity with the operation of an SDI system. Prior to their introduction to the SDI system, the subjects were asked to keep diaries for a two-week period of all documents of a professional nature that they read. At the end of this time, the subjects were interviewed regarding their reading and information gathering habits. Then the subjects were exposed for a 10-week period to the SDI system. On the last fourteen days of the experimental period, the subjects were again asked to keep diaries of their professional reading. At the conclusion of the experiment, the subjects were again interviewed. From the results, thirteen statements were hypothesized and the truth of these statements tested by a statistical analysis of the data collected.

Hypothesis 1: "SDI-1, at an average processing rate of twenty-three documents per week, does not affect the amount of time users read or the number of reading acts." This hypothesis was found to be acceptable. It appeared as though the SDI system caused a reallocation rather than an overall increase of reading time.

Hypothesis 2: "Assuming that the analyzed diary periods contain representative samples of the user's reading habits, SDI reallocates reading time toward SDI documents and away from other sources of information. Even though the participants may have tended to keep more careful records on SDI items read, this hypothesis is supported by the data.

Hypothesis 3: "It is not the lack of time to read that hinders users from keeping abreast of their area of work, but rather the lack of time to find relevant documents." Approximately 80% of the subjects expressed a willingness to look at five or more documents per week and 50% at ten or more documents per week. These results would tend to support th

hypothesis. The user accepts documentation from the system whether or not the information contained in that documentation is directly pertinent to a specific problem currently experienced.

Hypothesis 4: "Users prefer a two-stage SDI system as opposed to a single stage system." Subjects exposed to the one-stage SDI system acclimated to it, although they initially expressed preference for receiving abstracts first.

Hypothesis 5: "Users do not want to place a limit on the number of notifications received." Since the number of notices sent to each user was not high during the test period, the truth of this statement is not conclusive, even though two-thirds of the subjects indicated that they would not like to place a limit on the number of documents of which the SDI system notified them weekly.

Hypothesis 6: "Users want to be able to obtain personal hard copies of documents." Only 17% of the subjects were willing to receive a circulated copy. Two-thirds of the subjects wanted a personal copy.

Hypothesis 7: "Users are bothered by throwing away unwanted hard copies." 67% of the subjects said that throwing away documents that were of no interest or use bothered them. Of these, 96% would have preferred sending the unwanted documents back.

Hypothesis 8: "SDI conditions users to expect faster hard copy service." Those subjects who had to order documents considered six days a reasonable waiting period. Those subjects to which documents were automatically supplied considered twelve days a reasonable waiting period.

Hypothesis 9: "Users are exposed to useless documents in the course of their work, but they tolerate the present amount of trash." Approximately 40% of the subjects indicated that they received 5 or more documents per week that were of no interest but this quantity did not bother them.

Hypothesis 10: "The present library function is to provide document hard copies on request, not to refer a document to a user." The IBM library referred less than 1% of the documents read by the subjects to those subjects. This supports the hypothesis.

Hypothesis 11: "In a mixed research and development group, administrative and professional time spent reading is independent of profession age, managerial status, and degree level." For the subjects chosen, this hypothesis was true.

Hypothesis 12: "There is a significant time delay between the time users first order documents of interest and the time at which they actually read the document, even though there is little delay in receiving it." Under the conditions of the experiment, this hypothesis could not be adequately tested.

Hypothesis 13 deals with the effect on completeness of shortening the diary period and is thus only pertinent in the context of the authors' study.

The above conclusions are extremely valuable in designing an SDI system to meet specific objectives. No matter what objective the system seeks to meet, however, it seems clear that neither the number of documents read nor the length of time spent in reading will be materially influenced. It can be expected, though, that the direction of the users' reading habits will be significantly influenced. This factor must be considered when designing an SDI system to cover documentation not normally accessible to the user. A finding that affects the cost of the system considerably is the fact that titles alone provide a sufficient basis for determining the relevance of a complete document. A hypothesis never sufficiently tested, however, is that providing a complete abstract of considerable informative value results in a reduction in the number of hard copy documents ord[illegible].

H. EFFECT OF THE USER ON OTHER INFORMATION ACTIVITIES OF THE LIBRARY

The document supply activity of the library is certainly most readily affected by the users' response to an SDI system. The U. S. Army Electronics Command reports a four-fold increase in the number of documents ordered from Defense Documentation Center as a direct result of operating an SDI system. Before an SDI system is implemented, the document supply problem should be thoroughly examined to ascertain that the library is capable of implementing its stated policies on document supply. If the library states that its policy will be to supply documents on request in response to feedback notices from the users of the SDI system, then failure to uphold this policy would have an adverse effect on the system's reception by the users.

What has not been reported is the effect that the user has on library reference services. The question one might ask is, "Does the user of an SDI system increase his use of other library reference services to the extent that the library which operates such a system should increase its reference staff in anticipation of receiving more questions?" Certainly the increase in reference questions might be expected because of the library's more direct effect on the user.

Another area for investigation is a study of the change in attitude of the user toward the library as the result of being exposed to SDI. Does the user place more importance on the value of recorded information? Does the user assign a greater value to library services as the result of using an SDI system? The answers to these questions might be used to predict a change in the management support and use of the technial information staff.

IV. SDI SYSTEM DESIGN CONSIDERATIONS FOR ARMY ...

A review of the systems in current operation has demonstrated that the design of an SDI system rests primarily on the resources available to the library, such as number of personnel available for operation of the system and access to programs, digital computers, electronic accounting machines, and graphic reproduction equipment. Of secondary importance are the number of users served and the document input rate, since these really determine the number of personnel presently available for operating any information system. That is to say, the size of the library staff is usually proportional to the number of clients served and to the number of input documents processed. Design criteria for an SDI system applicable to most Army technical libraries is thus dependent on present technical library configurations studied with respect to personnel and access to ADP and graphic processing equipment.

A. ACCESS TO GRAPHIC PROCESSING EQUIPMENT

Almost every Army technical library has either direct or indirect access to graphic reproduction equipment. For the most part, direct access consists of a firm arrangement for using an office copying device at a nearby office. Some of the libraries provide photocopies of journal articles and selected pages from books on a regular service basis to their patrons. In the past few years, the widespread use of office copying equipment in the Government has made available a wide variety of office reproduction processes to the library. Since almost all of the libraries are located within a research complex that requires the use of office reproduction equipment in the course of preparing reports and maintenance of office files, this equipment is, in most cases, readily available to the library. Some of the larger libraries own or lease equipment which is under their direct control.

Since graphic reproduction equipment is necessary to support the publications effort of the research group which is supported by the Army Technical Library, this equipment is also available to the library. On a scheduled basis, most libraries are presently utilizing this equipment to produce multiple copies of accessions lists for distribution to its clientele. Most Army technical libraries have the use of offset printing equipment and facilities for making paper photo offset plates, although no library is large enough to sustain a self-supporting reproduction shop.

In most installations, offset printing and plate-making is performed by a service organization which supports an entire installation. Although some installations have a priority system for performing reproduction services, most installations run open shops on a first-come first-served basis so that library reproduction requirements can usually be satisfied on a periodic schedule.

B. ACCESS TO ADP EQUIPMENT

Since the Department of Defense is one of the largest customers for

ADP equipment, each of the Services is heavily endowed with programs, digital computer equipment, and punched card equipment. Many systems already use a digital computer to implement present library services such as retrospective information retrieval and the printing of custom prepared listings, and for preparing library records, such as catalog cards and journal subscription lists.

Restrictions on the use of data processing equipment are in terms of scheduling, total running time, and equipment configuration. Due to the wide latitude in the types of equipment configurations available in the Army and the tight scheduling of many installations, it is difficult to predict with any accuracy the exact configuration and probable scheduling for any one of the technical libraries. In view of this fact, the system to be designed should utilize a minimal equipment configuration which is likely to be available to most Army technical libraries and should make minimal demands in terms of total running time. The system should also be designed so that libraries currently preparing catalog cards on machine-readable media can use this same information in the automated SDI system. Some libraries, like the U. S. Army Biological Laboratory and the U. S. Army Electronics Command, already operate an automated SDI system using input in machine-readable form which has been prepared externally. The SDI systems design should allow these agencies to integrate other kinds of input into their present SDI system.

A minimal equipment configuration for an SDI system would consist of a digital computer, at least three magnetic tape transports, a highspeed printer, and punched card input/output equipment. From an analysis of an inventory of Army electronic data processing facilities, this minimal equipment configuration was found to be available to over 75% of the Army technical libraries who might use an SDI system. Because the system must also allow the remaining 25% of the Army technical libraries to achieve some SDI capability, the SDI system's design should allow for the substitution of EAM equipment functions in the ADP processing steps in the system.

The only function of the SDI system which is extremely desirable to implement by digital computer is the selection function. Although simple Boolean functions can be implemented by the optical coincidence technique (Peek-a-boo punched cards), the preparation of optical coincidence cards is time-consuming and limited to very simple Boolean expressions. But for a limited number of users and simple request types, optical coincidence can be used to implement the SDI match function.

For small installations, it would be desirable to maintain most files on punched cards rather than magnetic tape, because these can be modified most readily without making undue demands on a computer installation. For instance, an address file for 1000 or less users would be more practica to implement by punched card techniques, even though the same operation could be performed by digital computer.

The majority of libraries would require less than one-half hour of computer time per week to implement the SDI matching function. This time

would double, or even triple, if other automatic record-keeping and statistical programs were built into the SDI system. Automation of these auxiliary functions in the early stage of SDI implementation would also introduce problems in making the SDI system compatible with other manual or automatic procedures, because the operational requirements of the SDI system itself will change with time and experience and these changes cannot be anticipated beforehand. Thus, early programming of all functions would eventually lead to extensive re-programming in order to modify the system to accommodate unanticipated requirements.

C. PERSONNEL

Most Army technical libraries consider themselves understaffed to accomplish all the goals that they have set for themselves. Some libraries are presently operated by only a single person or a small group of people working part-time in information support activities. Thus it would not do to design an SDI system which required present personnel to devote full-time to the system's operation. Most of the work presently being performed by the library staff should be directly usable in support of the SDI system. For instance, present subject indexing, descriptive cataloging, and abstracting should be extrapolated to the system rather than requiring a separate and parallel input processing routine. Manual typing should be kept to a minimum and all processes that can possibly be conducted by external service support branches should be designed in that manner.

One method of minimizing typing would be to design an SDI system which could utilize directly all possible externally produced cataloging, indexing, and abstracting, such as is available from DDC, NASA, CAS, and other large information agencies which publish abstract journals.

D. VOCABULARY CONTROL

Of all the systems studied, those systems which operate from a controlled vocabulary obtain the highest relevance rates. Although auto-indexing can be accomplished on a digital computer as is done by IBM in its consolidated system, the relevance ratios obtained are decidedly lower than those obtained by systems working from a carefully controlled thesaurus of indexing terms. For systems which desire to utilize pre-indexed material, user profiles must be phrased in the same terms as the document input. This requires the operation of an essentially dual system: Original indexing is performed from an internal system's thesaurus from which user profile terms are also selected; external indexing is accepted without modification, but user profiles must be constructed from the externally produced thesaurus. Computer programs could be written to translate profiles composed of internal thesaurus words into external thesaurus words, but the desirability of writing such programs is questionable due to the high effort involved in record updating. It would be desirable to create a system which uses only one thesaurus and performs original indexing on all input, because of the complexity of running parallel systems and the updating problems which are incurred. This philosophy is also extended to the indexing of different types of material, such as books, journal articles, and technical reports.

At present, many Army technical libraries utilize the Library of Congress subject headings to subject index books, but use a modification of the DDC thesaurus or other external thesauri index technical reports. Again, a computer program could be written to translate from one to the other. But why not simplify matters by creating a system's thesaurus which incorporates the subject matter described by both subject headings and descriptors?

The acceptance of machine-readable input in large volume usually implies that the input is not closely tailored to the information requirements of the user. Therefore, the investment in original indexing of such input might not be warranted. This situation is the only one which clearly indicates the maintenance of two separate thesauri: one covering the externally obtained machine-readable input, and the other for original indexing of documents not represented on the machine-readable input.

E. PROCESSING WORKLOAD

Since some input items might never be selected for dissemination, SDI system should be devised in such a manner that investment in abstracting and preparation of machine-readable input should be minimized until it is ascertained that the input will be selected at least once. The degree of processing prior to the determination of a match against a user profile should be determined by experience with the overall percentage of documents which are selected for dissemination to at least one client. Since this factor cannot be determined until experience has been gained in operating an SDI system, the system should be devised in such a way that pre-processing routines can be modified easily during the early stages of systems implementation and operation. If, by experience, 95% or more of the input documents result in a match, then pre-processing of all input (including abstracting and descriptive cataloging) might be warranted. But if experience proves that only 50% of the input results in a match, then pre-processing should be delayed until after the match has been determined. Since original indexing must be performed on all documents not accompanied by acceptable index terms, the position of the document processing function in the system's flow diagram can only contribute to savings in original cataloging, abstracting, and reproduction master preparation.

As individual SDI systems develop within each Army technical library, there will be many incentives for trading SDI notices. To facilitate this exchange of notices, the SDI system should provide the widest possible latitude in the restrictions on format and content of the SDI notice. In recognition of the universality of the EAM cards in data processing, the user notice and response form should be designed around this format. The user notice should contain no restrictions on the type of information to be included for dissemination (i.e., journal articles, books, etc.) or on the document derivative data printed, since information content and format might be dictated by the special requirements of individual Army libraries. At a minimum, the user notice should contain a free field for punching a user profile number and a document number. These two items of information should be punched in the corresponding positions in the response form. There should also be no restrictions on the format of the

response form, except to designate the inclusion of the user profile number and the document number. These are the only restrictions that must be placed on user notice and response form for the purpose of facilitating exchange of SDI notices between individual Army technical libraries.

At some time in the future, the larger information resources, such as NASA, DDC, AEC, CAS and NLM, and others, might produce and make available unpunched EAM cards containing a graphic image of information to be disseminated. (NASA formerly did provide STAR entries on EAM cards for use in its own SDI system.) This might consist of what now constitutes a standard entry in the DDC TAB or NASA STAR. Each Army technical library could then perform an SDI match on document derivative data extracted from the notice itself. The resulting distribution list could be used to collate individual groups of SDI notices to be sent to each user.

The next chapter outlines a system which conforms to the above functional requirements for Army technical libraries. The design provides sufficient latitude so that any Army technical library can implement a detailed systems design to tailor the system to their individual requirements without going outside the basic framework provided.

V. SDI SYSTEMS DESIGN

This chapter presents the design for an SDI system generally applicable to Army technical libraries. The system is designed so that the following objectives will be attained:

(a) Each library will be able to determine the detailed effects of the implemented SDI system on its own operations and also on the reading habits of its technical clientele, and will be able to modify the system to streamline processing operations.

(b) It will allow investigation of competing philosophies in the design and operation of SDI systems in varied user environments, and the selection of options to increase relevancy of output.

This system will also possess several additional design characteristics, among them the operation of an SDI system at a reasonably low cost, and the participation of the library to an extent which will not curtail its normal functions.

In order to achieve very general capabilities for all Army library systems, the SDI system design has been divided into two major processing areas (see Figure V-1). These are areas of graphic processing and machine processing. Whenever a new item is selected by the library for SDI processing, inputs are prepared for both the graphic and machine portions of the SDI cycle. Separation of these inputs is essentially a simple clerical function.

The machine-processing segment of this system consists primarily of a search for matches with descriptors in user profiles, and the generation of printed instructions based on the results of this search. These instructions are later used to direct the generation of the appropriate number of copies of the desired abstracts, and their distribution to the appropriate users. The searching function is typically performed on descriptors assigned to documents and user profiles. However, it will be feasible for a given installation to base its search upon other information or descriptive terms; author name, place of origin, contract number or patent number are examples.

The graphic aspects of the SDI system are concerned primarily with the reproduction and distribution of a notice sufficiently descriptive of the document so that the user may make a decision as to whether or not he wishes to see the full copy of the document, or sufficiently informative so that the user will not need to see the whole document. Generally one would distribute the title and abstract of the article, but it is equally feasible to distribute the title only. Likewise, for short documents, distribution of the entire contents of the document may be feasible.

The selection of the means of reproduction of the document description depends upon the individual installation's access to reproduction

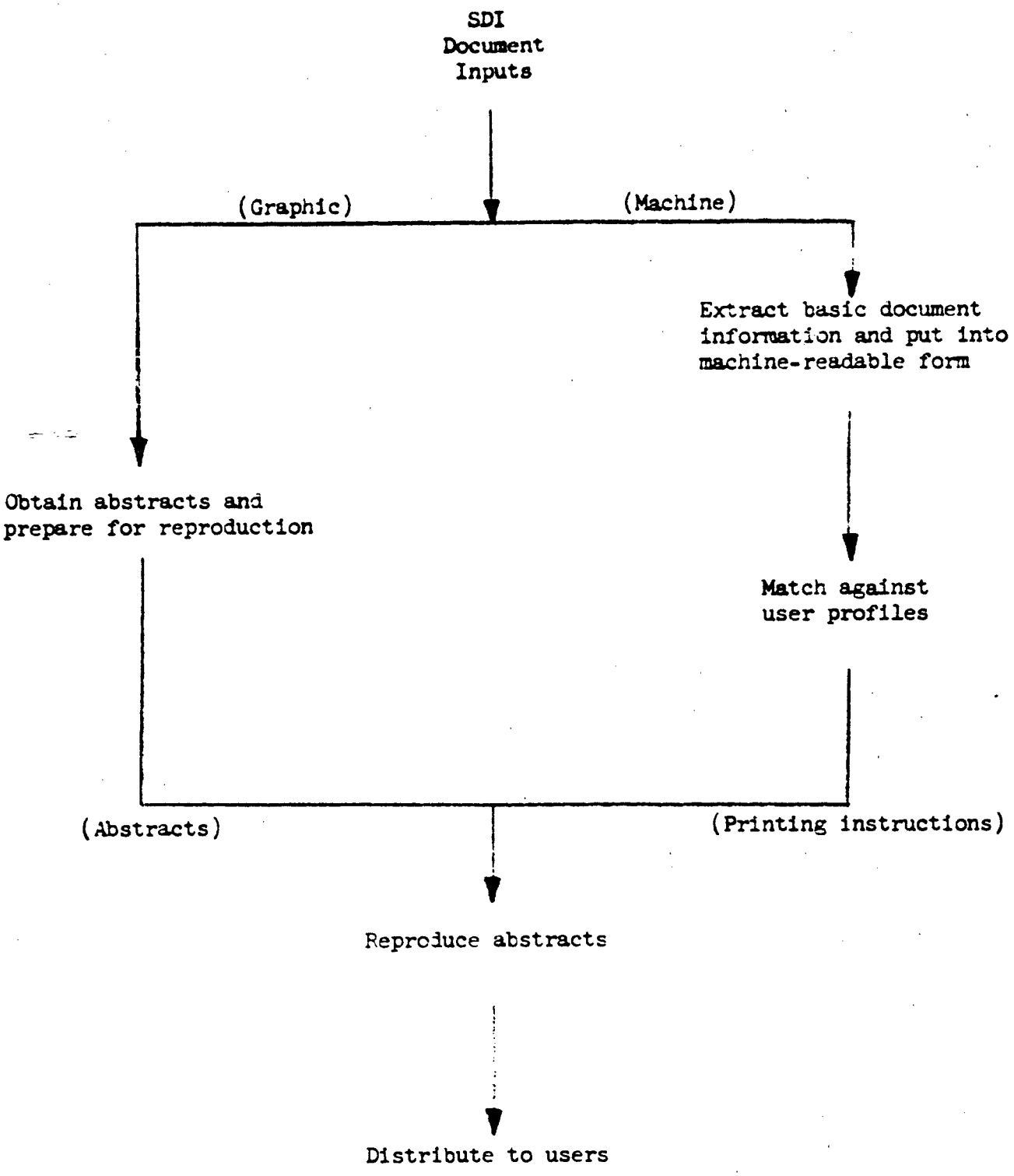


Figure V-1. The Concept of Graphic and Machine Processing in the SDI System

facilities, as well as the cost constraints which must be observed. It is possible to reduce all inputs to this system into a standard graphic format, from which it can be photocopied and reproduced onto the appropriate medium. On the other hand, the original abstract appearing with the text of the document as received by the library may be directly transcribed to the user notice by means of xerography or any other inexpensive office copying process.

Therefore, there are a large number of combinations of possible methods to transform the information input into user notices. The choice of a particular combination of methods rests with the individual installation; however, the basic resolution of the system into graphic and machine processes should be preserved.

The basic design of the SDI system is best presented in narrative form. The user is referred to the accompanying flow charts (Figures V-2 through V-4) throughout the following discussion of the SDI system design.

A. INPUT PREPARATION (See Figure V-2)

Information accepted by the SDI system consists of thre types of information derived from books, journal articles, technical reports, patents, and other kinds of documents. The three types of derivative information are:

(1) Descriptive cataloging information -- title, authors, corporate or journal source, etc.

(2) Index terms (descriptors, Uniterms, subject headings)

(3) Abstract or extract

The documents themselves may or may not be put into the SDI system, depending on whether the objective is to provide quick access to documents, or merely to provide an indication that such documents exist.

Document derivative information can be obtained by copying directly from the document itself, or from published abstract journals, such as the DDC Technical Abstract Journal, AEC Nuclear Science Abstracts, or CAS Chemical Abstracts (this same information may also be obtained on magnetic tape from these sources); or by performing original indexing from titles, abstracts, or documents, and creating original abstracts from documents.

Depending on whether original subject indexing and/or abstracting must be performed, information input can be classified into three types:

Type 1: Usable index terms and abstract are not available
Type 2: Usable abstract is available, but no index terms
Type 3: Usable abstract and index terms are both available

It is presumed that, at a minimum, descriptive cataloging information

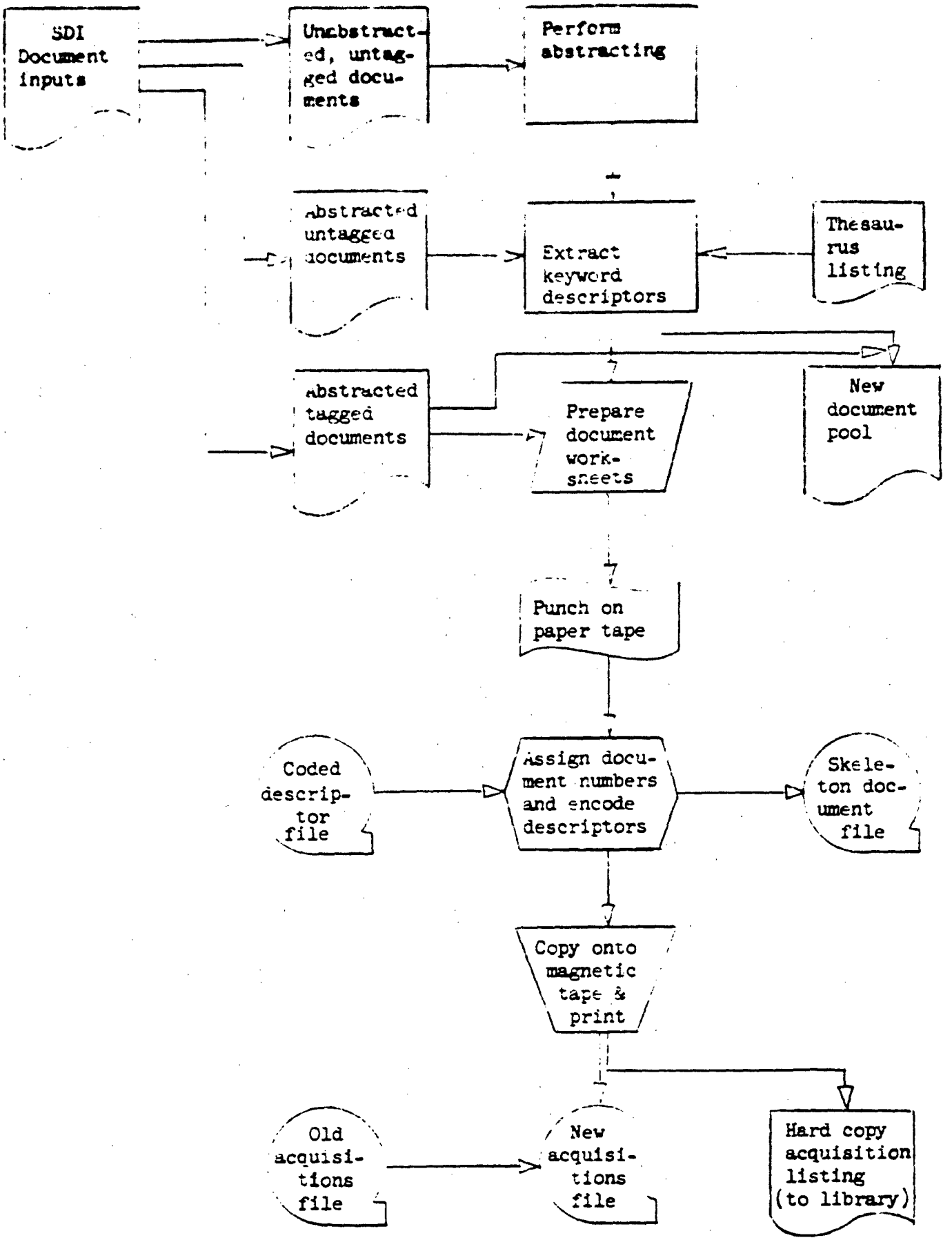


Figure V-2. SDI System Inputs and Input Processing

(title, authors, etc.) is available in all cases, since this is absolutely essential data.

In the systems description which follows, it is assumed that a very high percentage of the input items will yield at least one SDI notice. Later, alternatives will be considered in which this presumption is not true.

For every input item which is not already in machine-readable form, a worksheet is prepared that contains, at a minimum, descriptive cataloging information (including any document accession number assigned by the library). For Type 1 items, original indexing and abstracting is performed, and this information added to the worksheet. For Type 2 items, the abstract is copied and original indexing performed, and added to the worksheet. For Type 3 items, both abstract and index terms are copied to the worksheet. To this information is added a document availability code (i.e., library will supply copy on demand; library will order on demand; copy not available) and a document type code (i.e., journal article, technical report, etc.).

At this point, the worksheet might contain the information shown in Figure V-3. The information will be recorded on the form by office copying techniques, whenever possible, or by handwriting, if a suitable graphic image is not available for copying. The incoming documents are then physically placed in a special area, where they constitute the new document pool. The descriptive cataloging information, index terms and codes for document availability and type on the document worksheets just prepared are punched onto paper tape or cards in a special format which will allow the computer to distinguish among the various data elements in this input. When this information on all the document worksheets has been punched, proofed and corrected, it is taken to the computer facility for processing.

E. INPUT DATA REDUCTION

The first stage of computer processing of the input items consists primarily of putting them into improved machine-readable form, and of updating library history records to reflect the acquisition of new material. The input processor program accepts as its main input the punched paper tape or cards onto which all new items intended for SDI are recorded. This punched record is read into the computer at high speed, one logical record (i.e., one document description) at a time. The item is first assigned a unique processing number (which it retains for internal machine identification purposes) that is computed by taking the last such document number assigned incremented by one. Next, each descriptor from the input record is matched against a master file of such descriptors maintained on magnetic tape. For each descriptor, this master file is searched until an identical match is found. At this point the code number for that descriptor is inserted, along with the new processing number, into the record of the skeleton record file which is being created. After all such descriptors for a given document have been looked up and coded, the complete document record

DESCRIPTIVE CATALOGING INFORMATION:

Title	The precessing system of reference.
Author	Z. Karni
Publication date	1 May 1965
Pagination	7 p.
Source number	AFOSR 66-024
Accession no.	U-1356 (assigned by the library)
Order no.	AD 631031
Supplier	DDC, CFSTI $1.50

INDEX TERMS:

Precession; Differential equations; Equations of motion; Integration; Motion; Vector analysis; Momentum.

ABSTRACT:

The use of a precessing system of reference in the solution of inhomogeneous differential vector equations of the first and second order is demonstrated. (Author)

DOCUMENT AVAILABILITY AND TYPE:

Copy supplied, on demand (CODE 1): Journal article (CODE J)

Figure V-3. Sample Worksheet Document
Derivative Data

is written on the skeleton document file tape. Note that this tape will contain records which consist only of processing numbers and numerical codes for the descriptors pertaining to that document. No text is included in the skeleton file records.

Now all the information pertaining to the current document which was contained on the input record is transcribed onto magnetic tape in the form of a record on the new acquisitions file tape. This file contains all the pertinent information relating to every item which has been processed by the SDI system to date. Document information is stored on this tape in order of processing number assigned. In addition to recording this information on magnetic tape, a printed hard copy record of the information pertaining to the document is also written on the highspeed printer. Following this computer run, the listing thus generated may be taken to the library for updating of its acquisition records. The magnetic tape produced should be filed so that it may be updated at the next SDI run.

Summarizing the computer operations to this point, we find that we require the following inputs to the system:

1. A magnetic tape dictionary consisting of the descriptors in full English text and the numeric code assigned to each.

2. A magnetic tape file consisting of pertinent information on past items acquired for the SDI system.

3. The actual punched records containing descriptive information for new SDI acquisitions.

As outputs from this phase of the computer run we have the following:

1. A magnetic tape file consisting of skeleton records describing each of the input documents (i.e., processing number and codes for descriptors).

2. A magnetic tape file consisting of the old acquisitions file (an input to this run) updated with the descriptive information pertaining to new documents.

3. A hard copy listing of the information added to the new acquisitions file.

At this point all manual and computer-assisted preparation of input for the matching function has been completed.

C. MATCHING BY COMPUTER (See Figure V-4)

There is a logical division at this point between the preparation and reduction of input data for the SDI system, and beginning of the matching function. In practice, however, it is not anticipated that a physical

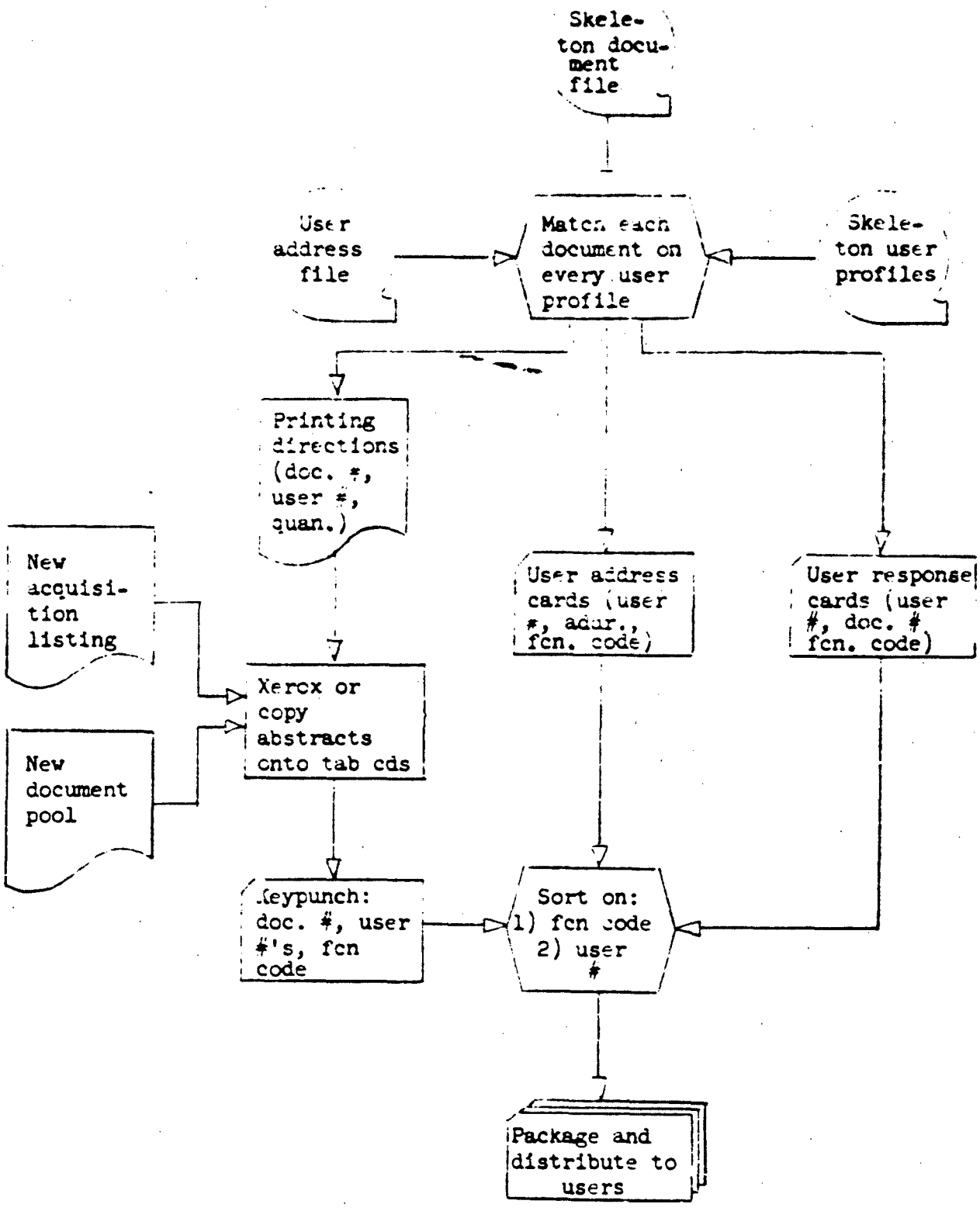


Figure V-4. SDI Matching Function and System Outputs

break in the running procedure will be encountered, since both phases are being accomplished with the assistance of the computer.

We now process (by means of the matching program) magnetic tape files containing the following information:

1. The skeleton document file which was prepared in the last phase of input preparation.

2. User profiles, also in skeleton form, containing for each user his unique user identification number and the numeric codes for the descriptors in his profile.

3. A tape containing the full name and office address of each user in the SDI system along with his unique identification number.

The matching program reads one record at a time from the skeleton document file described above. For each such record, all records in the skeleton user file are considered, and items from the two lists of coded descriptors are compared. If a match (according to the algorithm being used) is not found, the program merely proceeds to the next record on the skeleton document file.

If a match is found, a note is made of the user number whose profile matched with that of the document. If this is the first such match for this user, an address card is generated by searching the user address file (indexed by user number) and punching a card on the on-line card punch containing the user's full name and office address. In addition, a user response card is punched, consisting of the user number and processing number which constituted the match. Finally, after every user has been passed for the current document, a set of reproduction directions is printed on the on-line printer. These consist of the processing number, user number, and quantity of the document abstract required.

D. SYSTEM OUTPUTS

The computer processing of SDI information is now complete. The remainder of the tasks to be performed consist partly of graphic procedures (i.e., reproduction of document abstracts onto the desired paper forms), and partly of card processing.

We will first describe the graphic operations which are to be performed. The printing directions produced by the computer during the last phase of the run are taken to the reproduction area for processing. The contents of the document worksheets and new document pool are also made available for reproduction purposes. The reproduction room staff locates and reproduces the graphic image of the SDI notice for each document, either from the new document pool or the worksheets, through the following process:

1. The processing number and quantity of each graphic image to be produced are obtained from the computer-printed directions.

2. The processing number is looked up in the new acquisition listing produced by the computer in an earlier phase of operation described above.

3. With the full name and identification of the document at hand, the original document and its worksheet are located in the new document pool.

It now remains to reproduce the desired number of copies onto the appropriate paper medium. It is anticipated that this will be done in either of two ways: by Xerox dry-copy process, or by photographic reproduction onto mats for offset printing. The determination of which method should be used for a particular case is governed by two variables: the length of the abstract to be copied, and the number of reproduced copies required. If the abstract in its original size and format is sufficiently small to fit on one side of a standard IBM card, and not too many copies are required, then the Xerox process should be used. On the other hand, if the original abstract is too large, or for other reasons cannot be put on an IBM card, or if a great many copies are required, then the abstract should be photo-reduced to the required size, and an offset mat prepared for quantity printing.

After the required number of copies of each abstract have been prepared, the processing number and user number pertaining to each card are reproduced into these cards from the deck of computer-produced response cards. The resulting card deck is merged via a card sorter with the user address cards and user response cards generated previously, and the entire deck is sorted by user number.

The final results will thus be:

1. User address card for each user whose profile was found to match with at least one of the incoming documents.

2. A card containing a reproduced copy of the abstract of each matching document.

3. A response card containing the user number and document number for each matching document.

After all these cards have been sorted by user number, each package may be distributed to the addressed user via an internal mail system.

We will now examine in some further detail those data elements which are exogenous inputs to the system just described (i.e., which are not created by or maintained by the system itself).

The input of source items to the SDI system must be subject to

some filtering process in order to remove in advance any information which is obviously extraneous to the interest of the entire user group. One simple way of accomplishing this function is merely to accept all documents contained in a specified list of journals, periodicals, etc., and/or which are included under certain subheadings or subject indexes in abstract journals. A more effective pre-selection process might be obtained by having some professional or semi-professional staff members skim over the bulk of input items to the library and de-select those which obviously have no bearing on the interests of the user group. The choice of the method to accomplish this objective should be a decision variable left up to the library, based on its manpower capabilities.

The thesaurus listing and the coded descriptor file are merely hard copy and magnetic tape representations of the same information. Both contain code numbers for, and English text of, all the allowed descriptors in the official thesaurus. In the first case, the hardcopy listing is used by library personnel to index incoming documents; in the second case, the magnetic tape version is used by the digital computer to encode the descriptors into their numeric form.

Implied here is a master compilation of all allowed terms, under the specific management of the library. It is essential that additions, corrections, or deletions to this thesaurus be under the strictest supervision of the library. It is anticipated that this thesaurus will initially be compiled by the library staff, in consultation with the professional staff members who will participate in the system, and in consideration of existing thesauri such as those employed by DDC and other primary sources of abstract information. Following this initial compilation, the thesaurus will be updated or changed only in response to the wishes of the library management, which should in turn reflect the opinion of the user group. The physical thesaurus itself will be maintained on magnetic tape, and programs will be written for its updating, maintenance, and hardcopy printout.

The library acquisitions file is, as discussed above, merely a record of the "vital statistics" of all documents which have been routed through the SDI system. It is thought that a hardcopy record of this nature will assist the library staff in keeping records of various kinds relative to information which has been disseminated by means of the SDI system. The physical file itself will be created and maintained by subprograms within the SDI system pre-processor program.

The user address file and the skeleton user profile list are subsets of a more general list of information about SDI system users. This more general file consists of each user's complete name and office address, as well as a tabulation of all descriptors and weights in his profile. A maintenance program will be provided by which this master file can be reduced into more rapid-access form (i.e., skeleton-format tapes). The insertion and/or deletion of SDI users into/from this master file, as well as modification of individual profiles, must be provided for.

The matching algorithm which is to be used in the SDI system must be given careful consideration. It has been determined from study of the available literature that the utilization of "sophisticated" matching algorithms does not appear to contribute significantly to the relevance or usefulness of the system's products. Moreover, in all such systems, there are so many uncontrolled variables that it is impossible to ascertain whether or not the matching algorithm was responsible for correlation. Therefore, a design is provided by which a given installation may implement at least two different matching algorithms: one, a relatively simple sum of weighted matching descriptors, and the other involving Boolean expression evaluation. It is possible that experimentation with these different algorithms upon the user group may result in the emergence of one method or the other as the superior algorithm.

It is generally realized that a number of statistics concerning SDI system operation must be kept, if the degree of relevance of disseminated information and the "success" of the system generally are to be ascertained. The user response cards will constitute one measure of statistics collection; the document worksheets and acquisitions listing may be useful as others. However, the detailed implementation of a statistics-collecting process should wait until the system operation has begun, and has reached some kind of a "steady state". At such a time, it should become more evident just what records are required for an intelligent evaluation of system performance.

For the above system design, it was assumed that almost every input item would result in at least one notice. If the input information is not pre-selected, or pre-screened, this may not be the case. Many items may not yield notices, since the document described may be of interest to none of the SDI clients. To minimize the cost of input processing, the graphic processing and abstracting and descriptive cataloging functions may be delayed until after the documents (or derivative data) are indexed and the index terms matched against user profiles. Those items which are not selected for notification printing can be discarded. Selected items can be abstracted, descriptively cataloged, and graphically reproduced in the required number of copies. These procedural changes do not in any way affect the overall design of, or methods used by the system, yet they do provide a reduction in input processing costs.

APPENDIX A

SDI SYSTEMS SUMMARIES

At the onset of this project, an extensive bibliographic search was conducted to locate and obtain all the published literature pertinent to selective dissemination of information. Documentation obtained as the result of this search was divided into two parts: those documents which described the operation of a particular SDI system and those documents which examined a particular facet of SDI. On the basis of a preliminary analysis of the documents, a data collection form was developed and duplicated. A sample of the form, as completed by Mr. Warren B. McBirney for the Bureau of Reclamation System, is shown in Figure A-1.

Using the preprinted form as a guide, a complete analysis was performed on all available documentation. One form was completed for each system studied. At the conclusion of this analysis, personal visits were made to various installations and additional information was also collected by telephone contact and by mailing a copy of the preprinted form to the managers of selected systems. All information collected on each system is summarized in the following condensed data sheets. Wherever information is lacking, the corresponding heading is omitted. It should be noted that some of the summaries represent data collected from documentation (Code B), personal visit (Code V), correspondence (Code C), or telephone conversation (Code T), or any combination of these, as noted under the heading, "Information Source" on the condensed data sheets.

Although the systems operated by IBM at Chicago, Owego, Poughkeepsie, and Yorktown Heights Advanced Systems Development Division are now consolidated in the IBM system operating at the Thomas J. Watson Research Center, descriptions of these systems are included, because they contain operational data that is valuable in designing an SDI system. The following is a complete list of the systems studied:

American Cyanamid Co., Organic Chemicals Division, Bound Brook, New Jersey
Chemical Abstracts Service, Columbus, Ohio
Chemicals and Phosphates Limited, Haifa, Israel
Douglas Aircraft Co., Inc., Missiles and Space Systems Engineering, Santa Monica Division, Santa Monica, California
Dow Chemical Co., Midland, Michigan
Eli Lilly & Co., Indianapolis, Indiana
IBM Advanced Systems Development Division, Yorktown Heights, New York
IBM Data Systems Division, Poughkeepsie, New York
IBM Federal Systems Division, Space Guidance Center, Owego, New York
IBM Midwestern Region Sales Office, Chicago, Illinois

IBM Technical Information Retrieval Center, Thomas J. Watson Research Center, Yorktown Heights, New York
Institute for Scientific Information, Automatic Subject Citation Alert (ASCA), Philadelphia, Pennsylvania
Iowa State University, Ames Laboratory, Ames, Iowa
Mitre Corp., Bedford, Massachusetts
National Aeronautics and Space Administration, College Park, Maryland
Olin Mathieson Chemical Corp., E.R. Squibb & Sons Division, New Brunswick, New Jersey
Sandia Corporation, Livermore Laboratory, Technical Information Division, Livermore, California
U.S. Air Force Systems Command, Deputy for Foreign Technology, Wright-Patterson Air Force Base, Dayton, Ohio
U.S. Army Biological Laboratories, Fort Detrick, Maryland
U.S. Army Electronics Command, Information Office, Technical Information Division, Fort Monmouth, New Jersey
U.S. Department of the Interior, Bureau of Reclamation, Denver, Colorado
University of Pittsburgh, The Knowledge Availability Systems Center, Pittsburgh, Pennsylvania

Although this list of systems does not represent a complete listing of all automated SDI systems in current operation, these are representative of current trends in SDI systems design. It is recognized that many important systems unknown to the authors have been omitted from this study. Even after the publication of this report, however, the authors will continue to maintain an interest in SDI and will welcome receiving additional data and descriptions of systems.

SELECTIVE DISSEMINATION OF INFORMATION
SYSTEMS SURVEY

	I. General Information
A. Name of the system	Current Awareness Program--Selective Dissemination of Information
B. Location	Office of Engineering Reference, Bureau of Reclamation
Organization	
City	Denver
State	Colorado
C. Manager's name	Warren B. McBirney
D. Telephone number	303-233-3611, Ext. 6592
E. If the system is part of a larger, integrated information system, what other services or products are offered?	Manual and machine-based retrieval of technical information
F. What is the latest systems documentation?	
	II. Information Input (Please provide sample)
A. What types of publications constitute the input?	Preprints, reports, translations, journal articles not readily available to most technical personnel
B. What is the basis for screening or selecting the input?	Input oriented to agency's water resources development mission, with latitude to accommodate administration, management, writing, reporting, etc.
C. What levels of personnel perform the screening or selection?	Professional employees with bachelor's degrees or better in specific subject areas.

Figure A-1. Sample Data Collection Form

D. How are the publications abstracted, indexed, or otherwise characterized for machine input?	Inhouse-generated publications are informatively abstracted and indexed with descriptor with review by professional employees in Off of Engineering Reference; all others are indi atively abstracted and indexed with descripto within Office of Engineering Reference.
E. What device is used for machine input?	Abstracts, indexes, and bibliographic data prepared on punched paper tape for future co version to magnetic tape; presently, punched cards are used for machine input to magnetic tape.
F. What are the characteristics of the system thesaurus?	See sample page. Direct entry, for coordina indexing and retrieval; 3,200 terms with use references and descriptor grouping.
	III. User Group
A. What types of personnel make up the user group?	Engineers, scientists, and administrators in water resources engineering work.
B. How many users are there, and how many can there be?	Currently 2,200 users; unlimited as to population.
C. Where are the users located?	Throughout 17 Western States, Alaska, Hawaii, and Washington, D. C.
D. What are the subject interests of the users?	Planning, design, operation and maintenance and research on multipurpose water resourc projects.
	IV. User Profiles (Please provide sample)
A. How does the user submit his profile?	See sample profile form. Narrative stateme provided by user.
B. How is the user's profile indexed?	User's profile indexed with Reclamation's Th saurus of Descriptors by professional emplo in Office of Engineering Reference.
C. What limitations are placed on profile make-up?	Interest profile limited to 20 terms, four of which may be asterisked to indicate primary fields of interest.
D. What are the average profile specifications?	Prospective user prepares narrative statement without review by supervisor; most oriented about very specific technical interests requiring indexing in depth. Users require average of 17 descripto

	V. Selection Criteria
A. What basis is used for matching profiles against document descriptions?	See sheet on matching criteria.
B. What control does the user have over the matching criteria?	None directly; criteria set by study of number of matches and thereby dissemination volume.
	VI. Profile Maintenance
A. How is the user profile adjusted to reflect changing interests?	Profile may be changed upon request by user at any time; position changes are frequently noted and user so affected is invited to change his profile.
B. How is the user's address kept up to date?	By use of personnel roster and changes in office code number.
	VII. Statistics Computed (Please provide sample report)
A. What statistics are accumulated?	Statistics accumulated on SDI card dissemination, number of documents disseminated, and user interests in card information. See sample SDI report form.
B. How are they used in modifying the operating characteristics of the system?	Analyses compare current results of monthly disseminations with long-term statistics to determine whether serious departures occur. Special computer printouts are used to evaluate user's profile as an indicator of the need
	VIII. Products Disseminated for chan (Please provide sample)
A. What items of information are included in the notice to the user?	See sample SDI card, RECAP.
B. What provision for feedback is included with the notice?	SDI card has stub return features to request copy of document or to indicate interest only.

IX. Hardware and Software

A. What data processing equipment is used?	Honeywell H-800 and H-200, card punch equipment, and Flexowriter.
B. In what programming language are computer programs written?	COBOL (compiler) and ARGUS (machine processing)

X. Statistics Collected from Operation

A. % of notices resulting in:	
Positive response	47%
Negative response	-
No response	42%
Request for document	11%
B. Notices/User/1000 Documents	172
C. Machine time to:	Past statistics not meaningful since only 9 months' experience had with Honeywell equipment, and programs are still being adjusted for better efficiency. Recent run on 50-document input indicates that SDI card cost is approximately 13 cents.
Match 1000 documents	
Process-in 1000 documents	
Print one notice	

XI. Comments

SDI SYSTEM DESCRIPTIONS

American Cyanamid Company
Organic Chemicals Division
Boundbrook, New Jersey

INFORMATION SOURCE B

CURRENT STATUS Active

DOCUMENT INPUT

19 journals of general chemical nature. 9 journals were abstracted from cover to cover during a trial period. 10 journal titles were added in January 1966.

INDEXING AND ABSTRACTING

Each abstract contained an average of 150 words and was enriched by addition of an average of 10 descriptives taken from the whole text of the article.

USER POPULATION

A pilot run was conducted with thirty men from all areas of research and development who ranged in academic level from B.S. to Ph.D. As of June, 1966, 45 new subscribers were added.

USER PROFILES

Users were allowed to specify exact terms, route terms, and consecutive word terms. Each term was weighted with weights ranging from minus 9 to plus 9. Users could also specify MUST terms and NOT terms.

PROFILE MAINTENANCE

Profile changes were received via a response form printed out with the user notice. Users could specify three types of positive response and one type of negative response.

DOCUMENT/PROFILE MATCH STRATEGY

Terms in the user profile were matched against the entire abstract, descriptive cataloging information and descriptors for each document. Selection was based on comparing the algebraic sum of the weights of matching terms against a threshold stated in the user profile. If a MUST term was specified, the document was selected for notification regardless of weights. If a NOT term was present, the document was deselected regardless of weights.

EQUIPMENT IBM 1401 Computer

COMMENTS

American Cyanamid Company has discussed the possibility of utilizing machine-readable output from Chemical Abstract Service.

Chemical Abstract Service
Columbus, Ohio

INFORMATION SOURCE B and C

CURRENT STATUS Active

DOCUMENT INPUT

Chemical Abstract Service processes 690 current journals covering chemistry and chemical engineering and 500 journals covering the literature on the interface between organic chemistry and biochemistry, including effects of organic compounds on animal and plant systems, and in vitrochemical reactions of organic compounds of biochemical interest.

INDEXING AND ABSTRACTING

Document input to Chemical-Biological Activities includes title, summary, author names, molecular formulas and keywords. Input to Chemical Titles includes document title, author and source.

USER POPULATION Government agencies and industrial concerns

USER PROFILES

User profiles are specified in narrative style for individuals, groups, or whole organizations.

PROFILE MAINTENANCE Users may adjust profiles at any time.

DOCUMENT/PROFILE MATCH STRATEGY

Boolean search is performed using the functions AND, OR or NOR with a weighting factor.

USER NOTIFICATION AND RESPONSE FORM

Notification can be in the form of author indexes, bibliographies or Key-Word-In-Context indexes or any combination of the three.

EQUIPMENT IBM system 360/40.

COMMENTS

Chemical Abstract Service will supply one-time or continuing searches on the magnetic tapes it generates in the course of publishing Chemical Titles or Chemical-Biological Activities. It will also supply the tapes themselves. A whole family of SDI systems has grown from the use of CAS tapes. Chemical Abstract Service has also participated in the design and programming of SDI systems based on the use of its tapes.

Chemicals and Phosphates Limited
Haifa, Israel

INFORMATION SOURCE B

CURRENT STATUS Unknown

DOCUMENT INPUT Journal articles and books

INDEXING AND ABSTRACTING

Key words are selected from the American Institute of Chemical Engineers Thesaurus and assigned a fixed character code. One digit role code is added. A zero punch indicates term negation when used in conjunction with the single digit role code. Up to two codes with role indicators are punched on one card for each document.

USER POPULATION

About 100 professional people with widely diversified interests currently use the system.

USER PROFILES

Profile cards contain up to 12 codes per card. Terms are selected from the A.I.Ch.E. Thesaurus.

PROFILE MAINTENANCE

The user may indicate change of profile on the return stub of the notification.

DOCUMENT/PROFILE MATCH STRATEGY

Match of a single key word causes selection. Use of the role indicators limits output somewhat. If the document does not match any profile, a printed reference is sent to the technical information department.

DOCUMENT SUPPLY

Journal articles are routed automatically on the basis of standing orders and SDI selection. Thus, SDI notice for a journal article acts to key the user to look for articles in the current journal issue that will be circulated to him in the near future.

EQUIPMENT IBM 1401 4K core storage; no magnetic tapes

COMMENTS

This is the only system reporting the circulation of journals on the basis of SDI selection.

Douglas Aircraft Company, Incorporated
Missiles and Space Systems Engineering
Santa Monica Division
Santa Monica, California

INFORMATION SOURCE B

CURRENT STATUS Active

DOCUMENT INPUT

Technical reports only; no books, journal articles or others. 500 additions to their master reference file every two weeks.

INDEXING AND ABSTRACTING

The Uniterm method of subject indexing is used. Abstracting is not performed.

USER POPULATION

125 individuals and organizational units whose interests are research, engineering, testing, and production of air frames, missiles and space vehicles. Subject fields include astrophysics, biomedical and life sciences, electronics, flight mechanics, propulsion, and structures.

USER PROFILES

Users are interviewed by retrieval specialists and verbal descriptions of information required are translated to descriptors selected from a controlled source.

PROFILE MAINTENANCE

Individuals resubmit profiles if results are poor and on occasion submit minor revisions.

DOCUMENT/PROFILE MATCH STRATEGY

The user can specify the number of retrievals desired, what library division tape to be searched, date parameters and whether output is a card or a list. The sum of the weights of matching descriptors is used for selection. Profiles may also contain negative terms.

DOCUMENT SUPPLY Documents are supplied on request from the library.

STATISTICS ACCUMULATED

Report on deleted profiles, the number of profiles on record and for each profile, a list of the accession numbers selected, and the interest shown by the user.

USER NOTIFICATION AND RESPONSE FORM

The notification consists of a library card containing the accession number, corporate and personal authors, title, pages, date and descriptors. Response card contains options to indicate documents of interest -- notify when available; document of interest -- but not wanted; document of interest -- have seen copy; or -- document of no interest.

EQUIPMENT IBM 704 CDC 160A; IBM 7094 - 1401

RESULTS REPORTED

50 to 60% relevancy. Some profiles obtain 91% relevance. For an individual branch library, the total running time was 15 minutes every 2 weeks. Printing and assembling take about 1 hour. The bibliographic program takes 45 minutes on an IBM 7094 every 2 weeks.

Dow Chemical Company
Midland, Michigan

INFORMATION SOURCE B and V

CURRENT STATUS Active

DOCUMENT INPUT

Input to this system includes information contained in Chemical

Titles and Dow internal information. Tapes are purchased from Chemical Titles and hence cover the 600 to 700 journals covered therein. Additional material is being considered. No screening is done as far as individual items are concerned. Magnetic tapes are purchased with the material already on them.

INDEXING AND ABSTRACTING

The words in the titles of all articles and reports covered are selected. To this index are added journal and author names. Since titles are used, no vocabulary control is exercised.

USER POPULATION

Research scientists, chemists, biologists, mathematicians, physicists, sales personnel, physicians, virologists, marketing research personnel, computer personnel, etc., are all users of this service. There are at present about 200 active users and the system is set up to handle any number. Users are located throughout the U.S. at the various Dow locations. The subject interests of the users are varied but in general they include chemistry, physics, mathematics, biology, medicine and pharmaceuticals.

USER PROFILES

The user may submit his profile in several ways. He may do so during a personal interview in the course of which he is guided in his word selection or where he merely submits a list of words indicating the combinations for which he wishes to search. No indexing is done to user profiles. He must submit it in indexed form. Words are split to conform with the system employed by Chemical Titles. Users can use any words they wish. They may use as many words in one profile as they choose as well as authors' names and journal codes. An average profile includes about fifty terms but profiles range from 4 to 300 terms.

PROFILE MAINTENANCE

Any time a user wishes to make any additions, corrections or deletions to a profile, he submits them and they are made before the next search is undertaken. Any address change must come from the user and then it is handled in the same manner as any other profile change.

DOCUMENT/PROFILE MATCH STRATEGY

The weight-hit level concept is used. Each profile is assigned a hit level and then weights, both negative and positive, are assigned to the words so that the combinations of interest equal or exceed the hit level. Weights and hit level may be manipulated to effectively indicate the degree of similarity desired between documents and profile.

DOCUMENT SUPPLY

Since the library does not have copies of all documents, the responsibility of locating copies falls on the user.

STATISTICS ACCUMULATED

Statistics are accumulated on the following items: number of references printed as hits on each profile, number of hits that were of interest, of some interest, and of no interest. General information is also tabulated, such as average number of words per profile, average number of hits per person, etc. Figures showing percentage of hits that were of interest are used to determine if the system is operating correctly (a sudden drop for a number of individuals indicates something amiss). This figure is also used to determine if individuals should be encouraged to modify profiles and to determine if new sources of input are necessary.

USER NOTIFICATION AND RESPONSE FORM

A user receives with his notice the title, author, journal name, volume number, and inclusive pages of the article. The card is dated and there is a memo which indicates the issue of Chemical Titles in which the reference may be found. There is identifying information on the card, such as name, address and profile number, along with a hit total which is the total weight associated with the article. The card is a 4"x6" sized card which may be filed or used to request a copy of the article directly from the library. The incorporation of abstracts is being considered. A tear-off stub is affixed to each card on which user is to indicate if article is of interest. There is a space for the comments which may also be used for requesting profile changes.

EQUIPMENT

A Burroughs B-5500 Computer is used for the entire operation. All programs are written in Algol 60 with some sorting procedures written in Cobol.

RESULTS REPORTED

The entire group has an average interest figure of 58%. Those that have been on the system nine months or more average 70% of interest and about 86% respond regularly with feedback information. Out of an average of 400 input items, 85% of the users will retrieve an average of 39 items per person, although the range is from 1 to 360.

COMMENTS

The system is very satisfactory and has been readily accepted.

Eli Lilly and Company
Indianapolis, Indiana

INFORMATION SOURCE B

CURRENT STATUS Active

DOCUMENT INPUT

Magnetic tape records from Chemical Titles published by Chemical Abstracts Service.

INDEXING AND ABSTRACTING None is performed.

USER POPULATION 43 members of the Eli Lilly scientific staff

USER PROFILE

The individual served is asked to describe his scientific interests with a set of words that would, in his opinion, appear in the titles of papers. The set of words he provides is checked manually against the indexes of Chemical Titles. The editorial modifications (spellings, word roots, synonyms) are submitted to the requester for his approval and the edited list of words is incorporated into the request file. The requester can use author names, also. Average user profiles contain 38.2 terms per profile -- an average of 1.2 profiles per user.

PROFILE MAINTENANCE

Each individual may add or remove words from his request vocabulary any time he wishes.

DOCUMENT/PROFILE MATCH STRATEGY

Provision is made for simple inclusion, exclusion and negation.

STATISTICS ACCUMULATED

4 hours of 1401 time and 1.4 hours of 705 time were used for 55 profiles containing 2100 terms for a single issue of Chemical Titles which contains approximately 3200 titles.

USER NOTIFICATION AND RESPONSE FORM

When a match is achieved, the key word, journal code, requester, and his department number are written on magnetic tape. Authors and full title are added by table lookup. Journal codes are converted to accepted ACS abbreviations for journals. After October, 1964 each selected reference with all its attendant information was printed on a separate card and sets of cards were distributed. The participant keeps the cards of interest in his

card file and returns card stubs as an indication of his action.

EQUIPMENT

IBM 1401, 4K memory; IBM 705, 40K memory; IBM 729 Model 3 tape drive at 556 characters per inch tape density.

RESULTS REPORTED

There has not been unanimous support from the scientific staff. A few prefer the printed edition of Chemical Titles. Some have been enthusiastic about the idea of receiving selected titles, but feel the biochemical or biological interest is not covered adequately by Chemical Titles. Others who have been unwilling to put in the effort required to obtain a good key words vocabulary or have been unwilling to restrict the scope of their interest have been inundated by references. Some subjects fields are not as specific or as well defined as others, and this has created some difficulty in key word building for certain individuals.

COMMENTS

There is a decided difficulty in maintaining a balance between recall and relevancy ratios, when an uncontrolled vocabulary is used for indexing and searching.

IBM Advanced Systems Development Division
Yorktown Heights, New York

INFORMATION SOURCE B

CURRENT STATUS

Inactive -- Merged into IBM TIRC, T. J. Watson Research Center

DOCUMENT INPUT

30% internal reports, 25% internal patent disclosures, 44% derived from 45 professional and technical journals, 1% contributed articles from participants in the system itself. Input rate between 500 and 600 reports or articles per month.

INDEXING AND ABSTRACTING

Controlled vocabulary consisting of words or phrases up to 120 characters in length. (Former system used truncated single word descriptors.)

USER POPULATION

Over 1200 scientific, technical, and administrative personnel.

USER PROFILE

User can elect to prevent selection of notices for which hardcopy documents are not available. Profile consists of descriptors, multiple descriptor phrases, authors, corporate and journal sources, and contract numbers. User specifies MUST, NOT or MAY terms. User can specify absolute number of notices that will be acceptable. Separate profiles are maintained for each user interest. Profile descriptors are edited against a dictionary

PROFILE MAINTENANCE

Computer analysis used to monitor user and system performance. User submits profile changes on response card.

DOCUMENT/PROFILE MATCH STRATEGY

At every SDI run, random selections are made. NOT overrides MUST and MAY; MUST overrides MAY. All multiple-descriptor phrases are considered MUSTS and selection is made if at least 70% of descriptors match in a multiple-descriptor phrase (3 out of 4, 5 out of 7, etc.). MAY terms must match up to user stated threshold.

DOCUMENT SUPPLY

Hard copy supplied on request for 50% of input document as long as supply lasts. Copyrighted materials not duplicated.

STATISTICS ACCUMULATED

Profile descriptor with date added and/or deleted; number of times match occurred; records of all notices issued, with dates; type of user response; date document copy was sent (if requested).

USER NOTIFICATION AND RESPONSE FORM

2 IBM cards printed side by side by high-speed lineprinter. User notice includes 3-line title and descriptive cataloging information, 16-line abstract, and 2-lines of profile descriptors which match document descriptors. Notice printed in 3"x5" card format. Response form is Port-A-Punch card, pre-numbered and pre-punched with a serial transaction number. (User notice formerly printed by multilith.)

EQUIPMENT

IBM 7090, 32K memory, 8 tape drives; IBM 1401, 4K memory, 1 tape drive; card read/punch and highspeed lineprinter.

RESULTS REPORTED 70-75% relevancy (result of earlier system)

COMMENTS

To test desirability of the system, divisions were required to pay a subscription fee for SDI service. Number of users dropped from 1100 to 700. Of the 400 subscribers lost, 200 had been transferred to another division which operated its own SDI system.

IBM Data Systems Division
Poughkeepsie, New York

INFORMATION SOURCE B

CURRENT STATUS

Inactive -- Merged into IBM TIRC, T. J. Watson Research Center.

DOCUMENT INPUT

IBM formal technical documents, IBM patent disclosures, IBM interdivisional meeting minutes, IBM informal and progress reports, government reports obtained from DDC, NASA and other agencies, technical society reprints, cooperative exchange documents from universities and research organizations, and books from 7 IBM libraries. Investigating feasibility of adding 700 journals.

INDEXING AND ABSTRACTING

60-word abstract, descriptive cataloging information, and 14-27 applicable descriptors selected from 4500 terms extracted from the A.I.Ch.E. Thesaurus and 2000 IBM supplemental terms. (Up to 3 terms may be assigned which do not appear in the Thesaurus, but these are not used for SDI matching.) Professionals at two locations perform indexing.

USER POPULATION

450 engineers and scientists in DSD, plus others in two other IBM divisions (46 building locations).

USER PROFILE

Prepared from questionnaire and personal consultation. Terms selected from A.I.Ch.E. Thesaurus and IBM Supplement. Profile contains MUST, NOT, and IMPERATIVE terms plus hit level.

PROFILE MAINTENANCE

Feedback from user plus statistical analysis of response level and profile match level.

DOCUMENT/PROFILE MATCH STRATEGY

Hit level of percentage of document descriptors matching profile descriptors (I/D) IMPERATIVE (overrides I/D), MUST (selects, if I/D exceeds threshold) and NOT (overrides I/D).

DOCUMENT SUPPLY

24-hour hard copy supply (35,00 pages reproduced/month), loans of coyrighted documents, microfilm at DSD laboratories.

STATISTICS ACCUMULATED User responses

EQUIPMENT IBM 1401

RESULTS REPORTED 75-80% of notices are considered relevant

COMMENTS

High relevancy ratio might be computed by dividing total number of responses marked "relevant" by total number of responses.

IBM Federal Systems Division
Space Guidance Center
Owego, New York

INFORMATION SOURCE B

CURRENT STATUS Active

INDEXING AND ABSTRACTING

A professional cataloger performs abstracting, descriptive cataloging and indexing. Terms are selected from the DDC thesaurus or from the text. Text terms are added to an open-ended thesaurus.

USER POPULATION

100 individuals or department representatives.

USER PROFILE

Terms are selected from open-ended list or DDC thesaurus. User specifies hit level.

PROFILE MAINTENANCE Profile changes indicated on response form.

DOCUMENT/PROFILE MATCH STRATEGY

Selection is based on percentage of document descriptors matching profile descriptors which exceeds thresholds stated by user.

STATISTICS ACCUMULATED

List of descriptors and documents matched and also list of documents not matched.

USER NOTIFICATION AND RESPONSE FORM

Dual notice/response form, printed by computer on side-by-side IBM cards. Notice contains 3-line title, descriptors, descriptive cataloging information, and abstract (printing uses full IBM card). The response form is a Port-A-Punch card.

EQUIPMENT IBM 1401-7090

RESULTS REPORTED

Of 10,267 notices sent, 7,396 responses were received, of which 5,105 were marked of interest. Relevance rate 49.8%.

IBM Midwestern Region Sales Office
Chicago, Illinois

INFORMATION SOURCE B

CURRENT STATUS

Inactive -- Merged into IBM TIRC
T. J. Watson Research Center

DOCUMENT INPUT

Pertinent items selected by IBM specialists in the Midwestern Regional Office from approximately 150 publications. Each specialist is assigned 2 or 3 publications for review. The publications include the Wall Street Journal, Fortune, Sales Management, Harvard Business Review,

Communications of the ACM, American Documentation, Management Science, Control Engineering, Meat Packaging, etc. Many items are from newspapers.

INDEXING AND ABSTRACTING

Articles are extracted into less than 200 words by three secretari Specialists underline key sentences, mark significant paragraphs, or write evaluative comments. The secretary indexes and also adds terms from an IE thesaurus of 200 descriptors based on product and application marketing lines. One girl can scan an article, type the extract, and supply descrip tors for 20 articles per day. Additional index terms are assigned by comp scanning of the abstract and comparing abstract terms against a 600-word exclusion list and a dictionary of 2,200 words.

USER POPULATION

378 marketing personnel. Plans are to expand to 2,800 persons.

USER PROFILE

A user may have up to 35 profiles, each containing no more than 50 words of no more than 38 characters each. Most profiles contain no more than 20-30 words. Users average 2 profiles each. The user assigns a weig to each word. Truncated terms are accepted.

PROFILE MAINTENANCE Changes received on response form.

DOCUMENT/PROFILE MATCH STRATEGY

Weights ranging from minus 9 to plus 9 are assigned to each profil descriptor by the user. The user specifies a threshold which, if exceeded by the sum of the weights of profile descriptors matching document descrip tors, will cause selection.

DOCUMENT SUPPLY

11% of respondees requested full-size documents.

STATISTICS ACCUMULATED

For each notice sent, the user's name and location, notice number, abstract number, and matching descriptors. Plans are to add number of occurrences of user profile words that appeared in abstracts considered of interest and of no interest.

USER NOTIFICATION AND RESPONSE FORM

Notification contains final descriptive cataloging information, abstract, matching profile descriptors in 3x5" card format. Notification and response form printed side by side by high-speed lineprinter.

EQUIPMENT

IBM 1401, with four (4) IBM 729 2-tape drives and K Core memory.

RESULTS REPORTED

Of 26,300 response forms returned, 11% of interest - document requested, 48% of interest - document not requested, 7% of interest - have seen before, and 34% of no interest. System selects 5% of input for each user.

COMMENTS

.127 seconds needed to compare one average profile against one average abstract. Abstract is printed in 1.8 seconds. This is the first operational SDI system which indexes automatically from abstracts. Auto-indexing takes 16 seconds per abstract, using an exclusion list of 600 common words, or 60 seconds per abstract, using exclusion list plus dictionary of 2,200 words. Exclusion list alone indexes an item by an average of 41 key words. Combined exclusion list and dictionary produce 22 key words per item.

IBM Technical Information Retrieval Center
Thomas J. Watson Research Center
Yorktown Heights, New York

INFORMATION SOURCE B, C, and T

CURRENT STATUS

Active -- Absorbed 4 systems formerly located at Chicago, Poughkeepsie, Yorktown Heights, and Owego.

DOCUMENT INPUT

IBM research and engineering project files, IBM technical reports, IBM invention disclosures, IBM released publications, IBM suggestions, technical press log, non-IBM technical reports, selected U.S. patents, and abstract/index bulletins. Non-IBM reports are selected from many journal and magazine articles, government reports, and various university and college reports and theses.

INDEXING AND ABSTRACTING

Descriptive cataloging information and the complete author abstract are keypunched in upper and lower case, variable word-length format, and processed by computer to assure that the data is in machine-readable format and accurately represents the original data base. The processing uses the

computer for validity, sequence, and spelling checking. In spelling checking, the new words being processed are compared against a master list (10,000-15,000 words) of correctly spelled words. Any mismatch is printed out for human examination to determine if they are misspellings or are new words. These mismatches carry the document line and word information to permit keypunching corrections. Upon completion of the IBM 1401 processing, the binary coded decimal tapes are converted by an IBM 7090 into binary tapes for searching.

USER POPULATION

IBM scientists and engineers at all IBM installations. Three satellite operations at San Jose, California; Endicott, New York; and the World Trade Laboratory in La Gaude, France. 1,200 users are located in the U.S. and over 250 users are located in Europe.

USER PROFILE

Users submit information profiles on a special IBM form which contains blank space for the user name, man number, position or title, job code, department name, building and location, telephone number, IBM division, job data, and manager's approval. The user submits his profile in narrative form. Specialists at the information retrieval center code the textual profile into a computerized profile. Boolean logic, adjacent word logic, individual words (such as technological terms or authors' names), and MUST/NOT logic are used in coding the profile. Boolean logic (AND/OR) can be used with single words as, for example, (information or documents or data or literature) and (retrieval or retrieving or searching or search or searchers). Adjacent word logic allows retrieval of two-word sequences, such as information retrieval and variants such as (information or literature) (retrieval or searching or search). The latter would retrieve documents on information retrieval, information searching, information search, literature retrieval, literature search, and literature searching. Boolean logic can also be used with multiple-word terms. NOT can also be used with AND or OR. MUST can be used to specify selection on the basis of a single or multiple word term regardless of word search criteria.

PROFILE MAINTENANCE

The Port-A-Punch card feedback from the user, plus statistical analysis on notices sent, is used as a check on the validity of the notices. The user may submit a profile change on his own initiative or the user may be contacted by systems personnel who solicit a profile change because of adverse results indicated by user feedback or statistical controls.

DOCUMENT/PROFILE MATCH STRATEGY

The computer program looks for a match between single and multiple word terms in the user profile and terms in the narrative abstract. Matching terms are analyzed to determine whether they satisfy the single word and adjacent word logic specified in the profile. If a sufficient number

of logical statements is matched, the abstract is selected for dissemination. The user can specify the number of logical statements that must be matched.

DOCUMENT SUPPLY

16mm microfilm cartridges or reels are distributed to 25 library locations throughout the United States and foreign locations. These reels contain all IBM originated documents and those non-IBM documents whose reproduction is permitted by copyright and material that is releasable to foreign nations. Documents are filmed in entirety and are available to the user for his viewing and immediate reproduction of selected pages. The local library satisfies the hard copy requests.

STATISTICS ACCUMULATED

Statistics are accumulated for each of the major data bases and for each individual user and all users combined. The statistical program supplies the total notification sent out, number and percentage of response cards returned, number and percentage of interest (with a breakdown into document not needed, send copy, and have seen document before), number and percentage not relevant, and, in addition, a list of users who received no notification in the current period, any users who failed to return their response cards within a specified period, and those users whose "not relevant" response exceeded a predetermined percentage.

USER NOTIFICATION AND RESPONSE FORM

The user notice contains descriptive cataloging information, the author abstract, and the words in the user profile which matched in the author abstract. It also contains a note on the document availability, such as "not microfilmed." A response form is a Port-A-Punch card with punching positions available for abstracts of interest (document not needed, send copy of document, and have seen document before), abstract not relevant to my profile, and comments for address change.

EQUIPMENT

The system currently used is an IBM 7090-1401, but plans are under way to convert to an IBM 360 system.

RESULTS REPORTED

Statistics are reported for February through September of 1965. The table of statistics is reproduced below. It should be noted that the percent of relevance column at the right indicates the ratio of notices returned, marked relevant to the total number of notices returned. If relevancy is calculated, as is done in many systems, by calculating the ratio of notices marked relevant to the total number of notices sent, the percent relevancy for IBM, non-IBM, and invention disclosures would be respectively, 45.3%, 48.0%, and 41.7%.

SUMMARY OF CIS RESPONSE STATISTICS
February-September 1965

Data Base	Sent	Returned	Relevant	Invalid	Not Relevant	% Not Relevant	% Relevar
IBM	91,867	52,815	41,564	1,277	9,974	18.9	81.1
Non-IBM	43,472	27,500	20,854	541	6,105	22.2	77.8
Inv. Discl.	36,480	21,277	15,241	851	5,185	24.4	75.6
COMBINED	171,819	101,592	77,659	2,669	21,264	20.9	79.1

Institute for Scientific Information
Automatic Subject Citation Alert (ASCA)
Philadelphia, Pennsylvania

INFORMATION SOURCE B and C

CURRENT STATUS Active

DOCUMENT INPUT Journal Articles and U.S. Patents

INDEXING AND ABSTRACTING

Document input is descriptively cataloged. Patents are subject-indexed as well.

USER POPULATION

Government and commercial institutions, by subscription.

USER PROFILE

Six different types of profiles can be constructed.

Type 1 -- What journal article or U.S. Patent refers to document
Type 2 -- What journal article or U.S. Patent refers to a docum whose first author is X?
Type 3 -- What journal article or U.S. Patent is authored by X?
Type 4 -- What journal article or U.S. Patent was sponsored by organization X?
Type 5 -- What U.S. Patents were assigned to organization X?
Type 6 -- What U.S. Patents fall in Patent Office Main Class or class X?

PROFILE MAINTENANCE

Profile changes can be submitted at any time.

DOCUMENT/PROFILE MATCH STRATEGY

An exact match is required between the item designated "X" and the corresponding item in a U.S. Patent or journal article.

DOCUMENT SUPPLY

Documents are not supplied by the system.

USER NOTIFICATION AND RESPONSE FORM

The user notification consists of a preprinted form, headed with the name, address, and account number of the user, the reporting date and the number of journal articles and U.S. Patents processed. For each item identified in the search, the original question and the citing article are printed.

COMMENTS

This is one of the first commercial SDI systems.

Iowa State University
Ames Laboratory
Ames, Iowa

INFORMATION SOURCE B, C and T

CURRENT STATUS Active

DOCUMENT INPUT

Research and development study in the physical science fields; books on all phases of engineering, mathematics, and data processing; handbooks of tables and dictionaries.

INDEXING AND ABSTRACTING

Abstracts from Nuclear Science Abstracts were key-worded or indexed following the guide line prescribed by the European Atomic Energy Community (EURATOM) Thesaurus. The system will accept any document, abstract, key word, etc., in a KWIC or Science Citation Index Source format. The format and sequencing must follow the same restrictions required for KWIC indexing.

USER POPULATION

Nine subjects were tested in an initial pilot test.

USER PROFILE

Each profile contained approximately 25 words or word clusters. About 1/3 of the profiles used in the experiment were chosen by persons with experience in the subject area involved. The remainder was selected by persons not actually working in the particular subject area, but having a good concept of what people working in the area were interested in and were working on at the present time. Profile words or word clusters were each assigned an initial significance value. Profile words could be truncated as well as full English text words. Significance values (weights) range between 0 and 1-4 significant figures. The user may have as many as 99 individual profiles. A word cluster may consist of 2 to 6 words and each word can consist of any number of characters. Each profile is limited to 10,000 words or word clusters. Single words may be negated.

PROFILE MAINTENANCE

The significance value for each term was automatically incremented or decremented when positive or negative response was received. In addition, users could specify additions, deletions or modifications in the wording of their profiles. The value of the increment or decrement to each weight is determined by the original value of the weight according to a trigonometric function. Larger increments or decrements are added to or subtracted from a significance value of .5. Increments or decrements are less for significance values in the vicinity of .9 or .1. At the option of the system operator, no response is interpreted as a negative response, which results in a decrement to all descriptors in the user profile which resulted in the selection of that notice. For profile descriptors which did not result in the selection of a notice, a different but slower increment function operated.

DOCUMENT/PROFILE MATCH STRATEGY

Originally, the threshold for selection is set at .5. The significance values for matching descriptors are added by subtracting the product of two significance values from the sum of these significance values. If a third profile descriptor matches, the significance value of that descriptor is added to the previously accumulated sum according to the foregoing sum-minus-product formula. Thus, the summation formula represents the total probability of the union of two or more events.

STATISTICS ACCUMULATED

In the pilot test, statistics were accumulated on the number of runs required for a profile word to reach its equilibrium and the selection participation ratio (percentage of profile words participating in document selection).

USER NOTIFICATION AND RESPONSE FORM

The notification consists of a limited amount of the original document, plus the threshold value, significance value of the first eight words or word clusters which were used in the summation, the user's name and address, and user's profile number. The response form contained options for marking document of interest -- copy requested; document of interest -- not requested; impartial response -- don't adjust profile; document of no interest; user above is absent; and comments. A Port-A-Punch card was used as a response form.

EQUIPMENT IBM 7074, 10K memory, 12 tape drives

RESULTS REPORTED

Experimental results, using 9 scientists as subjects were reported as follows:

SDI Run No.	No. of Input Titles	No. of Notices Sent	Term Participation	Run Time (Min.)	% of Interest
1	1,197	311	8.8%	63	55%
2	1,198	151	---	42	65%
3	4,176	119	8.8%	--	59%
4	1,861	659	14.0%	--	52%

COMMENTS

This is the only system reported in which weights assigned to profile terms are automatically incremented or decremented on the basis of positive or negative response from the user. Results from operation of this system indicate that automatic modification of weights achieves reasonable relevancy of output. Profile words reach equilibrium after 1 to 5 dissemination cycles.

The Mitre Corporation
Bedford, Massachusetts

INFORMATION SOURCE V

CURRENT STATUS Active

DOCUMENT INPUT

The entire library input of contractual report literature.

INDEXING AND ABSTRACTING

The DDC Thesaurus is supplemented by Mitre identifiers.

USER POPULATION

This system disseminates to approximately 20 groups of profession and engineers. The clientele has wide interests in electrical engineerin communications, automatic data processing, information retrieval, etc.

USER PROFILE

The user either selects descriptors from the thesaurus himself, c states his information requirements to an information specialist who inde these statements with descriptors from the thesaurus. Approximately 35 d criptors are used per profile. Descriptors consist of single and multipl word terms.

PROFILE MAINTENANCE

Although the capacity for changing profiles is built into the com puter programs, operating personnel at present do not know how to accompl this.

DOCUMENT/PROFILE MATCH STRATEGY

A document is selected if at least one descriptor in the user pro file matches one of the descriptors attached to the document.

DOCUMENT SUPPLY

Document copies are available from the library on demand.

STATISTICS ACCUMULATED

No statistics are accumulated.

USER NOTIFICATION AND RESPONSE FORM

The user notification assumes the form of a classified accessions list which contains the date, department number, number of copies request department head; and for each document referenced, the corporate author, tle, personal author, library accession number and date of publication. Responses from clientele are received via Codaphone, located in the libra At periodic intervals, the comments made on the Codaphone are played back the appropriate action taken, which, for the most part, consists of deliv ing documents.

EQUIPMENT

IBM 1410, with magnetic tape units.

RESULTS REPORTED

50 to 100 document requests are received per month.

National Aeronautics and Space Administration
College Park, Maryland

INFORMATION SOURCE B and T

CURRENT STATUS Active

DOCUMENT INPUT

World-wide scientific and technical aerospace reports announced in the STAR and International Aerospace Abstracts, i.e., anything published on aerospace.

INDEXING AND ABSTRACTING

Documents are indexed and abstracted by specialists in the several fields which make up the broad aerospace categories. Documents are indexed from a controlled vocabulary consisting of single words and phrases.

USER POPULATION

700 scientists and engineers at 22 widely separated locations, who are active in the field of aerospace technology, materials, systems components, engineering and physics, chemistry, etc.

USER PROFILE

The user describes his interest in his own words, or phrases his interest in terms of the controlled vocabulary. The user can specify single and multiple word terms, terms logically AND'd together, MUST terms, and NOT terms. He can also include contract numbers. Profiles constructed by the user are computer checked against an allowed list of terms. If terms are not on the allowed list, they are printed out for review. Terms found in the dictionary are assigned a unique code.

PROFILE MAINTENANCE

Profiles are reviewed by NASA specialists. No mention of the method formally set up for handling user requests to change profiles was made. Mention was made, however, of contacting the author of inappropriate profiles by phone or letter.

DOCUMENT/PROFILE MATCH STRATEGY

MUST terms will cause selection over all other criteria. Two-to-three-terms phrases must match entirely. Three-to seven-term phrases need match only partially on any combination of terms in the phrase, i.e., a 6-term phrase requires match of any four terms. The user can also specify a fixed number or all terms in the profile to match. NOT terms can be AND'd together. A Boolean search strategy is used for matching all other terms which are not MUST or NOT terms.

DOCUMENT SUPPLY

Full size copies of documents will be supplied on demand.

STATISTICS ACCUMULATED

The average number of matches per person per 1,000 documents and the percent of notices relevant.

USER NOTIFICATION AND RESPONSE FORM

For each item disseminated, a continuous listing is made consisting of descriptive cataloging information and descriptors. The continuous form feed listing is published on a 3-part carbon set. The response form is incorporated in the right margin of the user notification. The user retains the top copy and sends the second and third copies to the library. The library keeps one copy and sends one copy to the NASA Document Facility. The response form has three columns for indicating document "of no interest," document "of interest" (requested or not requested). Formerly, each individual item was printed on a 3x5" card format and contained an image of the corresponding STAR entry. The primary difference between outputs is that STAR entries contain an abstract instead of descriptors.

EQUIPMENT IBM 7094

RESULTS REPORTED

50% to 60% relevancy. 40% did not respond at all. 15% to 18% requested documents. There are 20 to 25 notices per user per 1,000 documents scanned.

COMMENTS

This system is very similar to IBM's SDI-5. (See IBM, Yorktown Heights, New York).

Olin Mathieson Chemical Corporation
E. R. Squibb & Sons Division
New Brunswick, New Jersey

INFORMATION SOURCE B

CURRENT STATUS Unknown

DOCUMENT INPUT

Magnetic tape representing the information contained in Chemical Titles, published by CAS.

INDEXING AND ABSTRACTING

No indexing and abstracting is performed. Only the information on the tapes is used for dissemination.

USER POPULATION

More than 25 research scientists in key positions at the Olin Research Center, New Haven, Conn. Participants were organic and inorganic chemists, chemical engineers and information scientists, at levels up through research managers and directors.

USER PROFILE

A single user profile consisting of terms relevant to some of the company's research interests was submitted to CAS for matching against each issue of Chemical Titles.

USER NOTIFICATION AND RESPONSE FORM

Notification consists of a listing of the titles which contained the terms specified in the profile. The listings were divided according to request terms and given through the department heads to departments whose interests were pertinent to the terms. Evaluation questionnaires accompanied all transmittals.

COMMENTS

Participants at Olin pointed out that the convenience, appearance and legibility of the reference lists enhanced their utility. They felt that the capability of the service to inform them of even a relatively few papers of interest within the shortest possible time is of great value. Many of the papers retrieved were from journals not commonly read by American chemists because of language difficulties.

Sandia Corporation
Livermore Laboratory
Technical Information Division
Livermore, California

INFORMATION SOURCE B

CURRENT STATUS Unknown

DOCUMENT INPUT

Includes books, maps, technical reports, translations, patents, reprints, symposia, microfilm, periodicals, photographs and pamphlets. Technical reports constitute 75% of the input.

INDEXING AND ABSTRACTING

Documents are indexed and abstracted on a work sheet containing descriptive cataloging information, abstract and index terms.

USER POPULATION

All Livermore personnel are serviced by the system.

USER PROFILE

Profiles used specific, rather than general, index terms. Subject terms are selected from an authority established by the library.

COMMENTS

This system is a direct adaptation of SDI-3, as described by Sowarby in Reference No. 76 of Appendix C.

U.S. Air Force Systems Command
Deputy for Foreign Technology
Wright-Patterson Air Force Base
Dayton, Ohio

INFORMATION SOURCE B

CURRENT STATUS Active

DOCUMENT INPUT

Scientific and technical exploitation program abstracts, miscellaneous open-source literature, information intelligence reports, translations

(human and machine) of finished intelligence reports produced by contractors and Foreign Technology Division personnel. 8,000 to 10,000 documents are input per month.

INDEXING AND ABSTRACTING

Standard indexing and classification by descriptors and standard bibliographic information (including author, title, source, date, etc.) are prepared on a Flexowriter. Descriptors are selected from a very carefully controlled thesaurus containing terms of 1-, 2- or 3-word expressions.

USER POPULATION

Users are groups rather than individuals. About 70 user groups are active and are spread widely throughout the United States. They show a diversity of missions, staffing, and subject interest.

USER PROFILE

User profiles are reviewed quarterly. CIRC prints user profile with number of times each term caused selection. The list is hand-carried by an analyst who reviews it with the users in the user group. Profiles are indexed by descriptors selected from the CIRC Thesaurus.

DOCUMENT/PROFILE MATCH STRATEGY

There is a clip level in the program -- a little formula -- by which they can vary the machine criteria which determines what shall be sent to a specific customer.

DOCUMENT SUPPLY

Documents are supplied on request, or are automatically disseminated.

STATISTICS ACCUMULATED

The number of hits per profile term are used in conjunction with reviewing user profiles.

USER NOTIFICATION AND RESPONSE FORM

At the option of the user, one or any combination of three levels of dissemination can be elected. One is header information, another is an abstract, and the third is the full document. The header contains accession number, date, country, subject codes, title, group of terms, author and source. An asterisk indicates document terms that match profile terms.

EQUIPMENT

IBM 7094, IBM 1401-1403 with 17 tape drives and two data channels.

COMMENTS

This is one of the few systems that disseminates three levels of information, all the way from title only, to full document.

U. S. Army Biological Laboratory
Fort Detrick, Maryland

INFORMATION SOURCE B, V and C

CURRENT STATUS Active

DOCUMENT INPUT

Magnetic tapes from the National Library of Medicine Monthly Announcement Bulletin identifying U. S. and foreign-produced journals and monographs indexed with appropriate medical subject headings.

INDEXING AND ABSTRACTING

No indexing or abstracting is performed. The MEDLARS tapes are prescreened by the National Library of Medicine by using a composite profile for the entire laboratory. A sub-set of the complete tape is then delivered to the laboratory. The subject indexing authority for MEDLARS tape is the Medical Subject Headings published by NLM.

USER POPULATION

50 scientists in the biological, biochemical, bacteriological and biophysical fields.

USER PROFILE

Users are initially interviewed for one-half hour to eight hours to obtain a verbal statement of interests. On the average, profiles contain 30 terms and an average of 4 profiles describe a user's interest. Profile terms are selected from Medical Subject Headings. Each profile term is assigned a weight ranging from minus 9 to plus 9. The user may also specify MUST and NOT terms.

PROFILE MAINTENANCE

Both terms and weights can be adjusted at the option of the user.

DOCUMENT/PROFILE MATCH STRATEGY

A system of weights summed and matched against a threshold is used

as a criterion for selection. MUST and NOT constraints may also be added. The way in which weights are used, however, indicates that a Boolean search strategy is actually being implemented. In each profile a few terms are weighted very heavily, but the majority of terms are weighted very low. Because of the extremes used in weighting terms, documents will be selected only if a highly-weighted term is coupled with a low-weighted term.

DOCUMENT SUPPLY

Some of the journal articles cited are available in the library, but the library is not responsible for supplying all documents.

STATISTICS ACCUMULATED

A compilation of the participants, number of notifications, responses and type of responses in both the current period and the one preceeding. This information is used to assist in the improvement of interest profiles and to record systems effectiveness. Profiles are reviewed when a number of notifications are relatively low or high.

USER NOTIFICATION AND RESPONSE FORM

The user notification contains the document title, source, authors and date of publication. The response form is printed on the user notice and contains four response options, three of which indicate document of interest.

EQUIPMENT

A Univac solid-state 90 system. Central processor with 1280 words of core storage, 2 magnetic tape drums with 8800 word storage. Card read-punch, 150 cards per minute. High-speed card reader, 450 cards per minute. High-speed printer, 600 lines per minute. Tape drives, Randex drum file with 24×10^6 digit storage.

RESULTS REPORTED

Relevancy obtained averages 50% with a range from 40 to 80%.

U. S. Army Electronic Command
Information Office
Technical Information Division
Fort Monmouth, New Jersey

INFORMATION SOURCE V, C and T

CURRENT STATUS Active

DOCUMENT INPUT — Technical reports from the DDC collection.

INDEXING AND ABSTRACTING

In a pilot test, entries were selected directly from the Technical Abstract Bulletin. These entries consisted of descriptive cataloging information, an abstract, and descriptors. Entries were used without modification.

USER POPULATION — 82 engineers and scientists at Fort Monmouth.

USER PROFILE

Profiles were constructed by selecting terms from the DDC Thesaurus.

PROFILE MAINTENANCE

Users are allowed to submit profile changes verbally or in writing.

DOCUMENT/PROFILE MATCH STRATEGY

During the course of an 18 month experiment, descriptors from all profiles were merged and a search made through a biweekly subject index to the Technical Abstract Bulletin for the occurrence of each descriptor in the merged profiles. When the descriptor was found in the index, a notation was made of the document indexed by that descriptor, and the users who had specified that particular descriptor in his profile. After eliminating duplicate occurrences of documents, one TAB entry was abstracted for each document selected. The profile numbers were posted directly on the entry and the entry was duplicated in the required number of copies. The copies of the TAB entry were then sorted by profile number and stapled together in a booklet.

DOCUMENT SUPPLY

On the basis of response, documents were ordered from DDC.

STATISTICS ACCUMULATED

Average number of abstracts per booklet, average number of interest, average number of documents ordered.

USER NOTIFICATION AND RESPONSE FORM

The user notification consisted of the stapled booklet, DDC Abstract Entries. The response form consisted of a cover sheet for listing the AD numbers of pertinent documents.

EQUIPMENT — No mechanized equipment needed.

RESULTS REPORTED

Average number of abstracts per booklet -- 25; average number of interests -- 70%; average number ordered -- 15%. Percentage increase in the use of DDC over prior 18 months -- 400%.

COMMENTS

This system is a prime example of the success to be obtained with limited equipment and minimum personnel sources. On the basis of the promising results obtained, computer programs have been flow charted for processing magnetic tapes generated by DDC. Negotiations are being concluded for the release of the DDC tapes on which the TAB entries are originally generated.

U. S. Department of the Interior
Bureau of Reclamation
Denver, Colorado

INFORMATION SOURCE B, C and T

CURRENT STATUS Active

Note: For details see Figure A-1, sample data collection form completed by Warren B. McBirney.

University of Pittsburgh
The Knowledge Availability Systems Center
Pittsburgh, Pennsylvania

INFORMATION SOURCE C

CURRENT STATUS Active

DOCUMENT INPUT

Aerospace reports from the following sources:

NASA Centers such as Lewis, Langley, Goddard, Marshall, Ames, etc.
NASA contractors, such as Westinghouse, RCA, General Electric, Battelle Memorial Institute, etc.
Government agencies, such as Atomic Energy Commission and Department of Defense.

Nation-wide Government-sponsored university research reports.
Technical society transactions.

As of June, 1966, over 200,000 documents were available for SDI and documents were being added at a rate of 6000 per month.

USER POPULATION

Corporate subscribers. 59 research and industrial firms are listed as subscribers.

USER PROFILE

Profiles for individual groups are formulated in consultation with a senior instructional staff member of the School of Engineering.

DOCUMENT SUPPLY Documents are supplied on request.

COMMENTS

This is the first known commercial use made of externally produced tapes. Input of the system consists of magnetic tapes received from Washington, D.C., presumably from the National Aeronautics and Space Administration.

APPENDIX B

BIBLIOGRAPHY OF SDI SYSTEMS CLASSIFIED BY SYSTEM LOCATION

American Cyanamid Co.
Organic Chemicals Division
Bound Brook, New Jersey

"Computer to Match Chemical Abstracts with User's Keyword Profile," Communications of the ACM, v. 9, n. 5 (May 1966) p. 401.

Gallagher, Joan. "Experience in the Introduction of an SDI System." Paper presented at Division of Chemical Literature, American Chemical Society 151st National Meeting, Pittsburgh, Pa., March 22-24, 1966.

Chemical Abstracts Service
Columbus, Ohio

"Computer to Match Chemical Abstracts with User's Keyword Profile," Communications of the ACM, v. 9, n. 5 (May 1966) p. 401.

Freeman, Robert R. "Automatic Retrieval and Selective Dissemination of References from Chemical Titles: Improving the Selection Process." In: Automation and Scientific Communication, Washington, D.C., 26th Annual Meeting of the American Documentation Institute, 6-11 October 1963.

Freeman, Robert R., et al. "Automatic Preparation of Selected Title Lists for Current Awareness Services and as Annual Summaries," 144th National Meeting of the American Chemical Society, Los Angeles, Calif., April 1963.

Chemicals and Phosphates, Limited
Haifa, Israel

Ofer, K.D. "SIDAR: Selective Information Dissemination and Retrieval," Journal of Chemical Documentation, v. 4, n. 1 (January 1964) p. 54-55.

Douglas Aircraft Company, Inc.
Missiles and Space Systems Engineering
Santa Monica Division
Santa Monica, California

Bunnow, L.R. Study of and Proposal for a Mechanized Information Retrieval System for the Missile and Space Systems Engineering Library. Santa Monica, Calif., Douglas Aircraft Company, Inc., May 1960. (SM-37418).

Koriagin, G.W. Library Information Retrieval Program. Douglas Aircraft Company, Inc., Engineering Paper No. 1269, January 19

Koriagin, G.W. and L.R. Bunnow. Mechanized Information Retrieval tem for Douglas Aircraft Company, Inc. Status Report. Dougla Aircraft Company, Inc., Report 39167, January 1962.

Williams, A.S. "Historical Development and Present Status - Dougl Aircraft Company Computerized Library Program." In: Working W Semi-Automatic Documentation Systems, Warrenton, Va., AFOSR/SI Airlie Foundation, May 1965, p. 41-54. (AD 620-360).

Dow Chemical Company
Midland, Michigan

Bowman, C.M. "Engineers' Personal Files and Internal Information in an Industrial Organization." Paper presented at the Symposium on Changing Aspects and New Services of Engineering Information: Fifty-eighth National Meeting, Philadelphia, Pa., December 5-9, 1965, American Institute of Chemical Engineers.

Institute for Scientific Information
Automatic Subject Citation Alert (ASCA)
Philadelphia, Pennsylvania

Sher, Irving and Eugene Garfield. "ASCA: An Automatic Current Alerting System Based on Citation Indexing." In: Congress of the International Federation for Documentation (FID), Washington, D.C., 10-15 October 1965.

IBM Advanced Systems Development Division
Yorktown Heights, New York

Barnes, A.B., A.A. Briggs, J. Gauss, and A. Resnick. "SDI-5, An Advanced System for Selective Dissemination of Information," ACM Proceedings of the 19th National Conference, 1964.

Brandenberg, W., H.C. Fallon, C.B. Hensley, T.R. Savage, and A.J. Sowarby. Selective Dissemination of Information, SDI-2 System. Yorktown Heights, N.Y., IBM Corporation, Advanced System Development Division, 1961.

Bushnell, R.C. Selective Dissemination of Information (SDI) Program Package for the IBM 1401. Akron, Ohio, IBM Corporation, Data Processing Division, 1962.

Hensley, C.B. Design Developments in Information Management Through Selective Dissemination and Retrieval System. Preprints of papers presented at the 16th National Meeting of the Association for Computing Machinery, 1961.

Hensley, C.B., T.R. Savage, A.J. Sowarby, and A. Resnick. "Selective Dissemination - Report on a Pilot Study - SDI-1 System." IBM-ASDD, Yorktown Heights, N.Y., Report 17-C39, January 1961, 44p., illus.

Hensley, C.B., A. Resnick, T.R. Savage, and A.J. Sowarby. Preliminary Report on the Selective Dissemination of Information System. Yorktown Heights, N.Y., IBM Corporation, Advanced Systems Development Division, 1959.

Hensley, C.B., T.R. Savage, A.J. Sowarby, and A. Resnick. "Selective Dissemination of Information - A New Approach to Effective Communication." In: The Coming of Age of Information Technology, Volume VI, V. Slamecka, ed., Bethesda, Md., Documentation, Inc., 1966.

IBM General Information Manual - Selective Dissemination of Information. San Jose, Calif., IBM Corporation, Product Publications Department, 1962.

Luhn, Hans Peter. "A Business Intelligence System," IBM Journal of Research and Development, v. 2 (October 1958) p. 314-319.

Luhn, Hans Peter. "Selective Dissemination of New Scientific Information with the Aid of Electronic Processing Equipment." Presented at the 126th Meeting of the American Association for the Advancement of Science, Chicago, Ill., December 29, 1959. (IBM, Yorktown Heights, N.Y., p. 225-1488).

Resnick, A. "Progress Report on IBM's Selective Dissemination of Information - SDI-4 System IBM 7090 - 1401 Data Processing System." In: Automation and Scientific Communication, Washington, D.C., 26th Annual Meeting of the American Documentation Institute, 6-11 October 1963.

Sowarby, A.J., et al. The SDI-3 System for the IBM 1401 Data Processing System. Yorktown Heights, N.Y., IBM Advanced Systems Development Division, 1962. (Report No. 10.3.04).

IBM Data Systems Division
Poughkeepsie, New York

White, Herbert S. "The IBM DSD Technical Information Center: A Total Operating System Approach Combining Traditional Library Features and Mechanized Computer Processing." In: Automation and Scientific Communication, Washington, D.C., 26th Annual Meeting of the American Documentation Institute, 6-11 October 1963.

IBM Federal Systems Division
Space Guidance Center
Owego, New York

Balz, Charles F. and Richard H. Stanwood. "Literature Dissemination and Retrieval Using the MERGE System." In: Automation and Scientific Communication, Washington, D.C., 26th Annual Meeting of the American Documentation Institute, 6-11 October 1963.

Benjamin, R., S.D. Miller, and E.S. Rowland. Selective Dissemination, (SDI). Owego, N.Y., IBM Corporation, Space Guidance Center, November 15, 1962.

Benjamin, R., S.D. Miller, and E.S. Rowland. Selective Dissemination, IBM 7090 Program, Owego, N.Y., IBM Corporation, Federal Systems Division, October 30, 1961.

Benjamin, R., S.D. Miller, and E. Scofield. Technical Report Selective Dissemination (SDI), Owego, N.Y., IBM Corporation, Federal Systems Division, October 30, 1961.

IBM Midwestern Region Sales Office
Chicago, Illinois

IBM Application Program H20-0053-1, Selective Dissemination of Information System for the IBM 1401 Version I (1401-CR-01X) Manual 1: Application Description.

IBM Application Program H20-0054-1, Selective Dissemination of Information System for the IBM 1401 Version I (1401-CR-01X) Manual 2: User Manual.

IBM Application Program H20-0055-1, Selective Dissemination of Information System for the IBM 1401 Version I (1401-CR-01X) Manual 3: Operator's Manual.

Kraft, Donald H. "Application of IBM Equipment to Library Mechanization, Keyword-in-Context (KWIC) Indexing and the Selective Dissemination of Information (SDI)." In: Proceedings of the 1963 Clinic on Library Applications of Data Processing, Champaign, Ill., April 28-May 1, 1963. Illini Union Bookstore, 1964, p. 133-156.

Kraft, Donald H. "An Operational Selective Dissemination of Information (SDI) System for Technical and Non-Technical Personnel Using Automatic Indexing Techniques." In: Automation and Scientific Communication, Washington, D.C., 26th Annual Meeting of the American Documentation Institute, 6-11 October 1963.

Kraft, Donald H. "Data Processing Equipment for Library Use in Clerical Tasks and Dissemination of Information." Illinois Libraries, v. 44, n. 9 (November 1962) p. 587-592.

IBM Technical Information Retrieval Center
Thomas J. Watson Research Center
Yorktown Heights, New York

Magnino, J.J. "CIS - A Computerized Normal Text Current Awareness Technique." Yorktown Heights, N.Y., IBM Technical Information Retrieval Center, Thomas J. Watson Research Center, November 1965, Technical Report No. ITIRC-006.

Magnino, Joseph J., Jr. "Textual Information Retrieval and the IBM Suggestion Plan." In: Congress of International Federation for Documentation (FID), Washington, D.C., 10-15 October 1965.

Merritt, Allen C. and Paul J. Nelson. The Engineer-Scientist and an Information Retrieval System. Yorktown Heights, N.Y., IBM Technical Information Retrieval Center, Thomas J. Watson Research Center, August 1966, Technical Report No. ITIRC-003.

Merritt, C. Allen. "An Operating System: The IBM Technical Information Retrieval Center." In: Working with Semi-Automatic Documentation Systems, Warrenton, Va., AFOSR/SDC/Airlie Foundation, May 1965, p. 32-39. (AD 620-360).

Iowa State University
Ames Laboratory
Ames, Iowa

Anderson, R.R., D.R. Fitzwater, and C.R. Sage. "The Stability of Dynamic Feedback in an SDI System." In: Proceedings of the 19th International Guide Meeting, October 15, 1964,

Sage, C.R., R.R. Anderson, and C.H. Cameron. Ames Laboratory SDI Systems Reference Manual, 1964. (U.S.A.E.C. Report IS-940).

Sage, C.R., R.R. Anderson, and D.R. Fitzwater. "Adaptive Information Dissemination," American Documentation, v. 16, n. 3 (July 1965) p. 185-200.

Sage, C.R. and D.R. Fitzwater. Ames Laboratory Selective Dissemination of Information Computer System. 1963. (U.S.A.E.C. Report IS-751).

Sage, C.R. and D.R. Fitzwater. "Operational Results of an Adaptive SDI System." In: Proceedings of the 19th International Guide Meeting, October 15, 1964.

Eli Lilly & Company
Indianapolis, Indiana

Freeman, Robert R., et al. "Automatic Preparation of Selected Title Lists for Current Awareness Services and as Annual Summaries," 144th National Meeting of the American Chemical Society, Los Angeles, Calif., April 1963.

Rice, C.N. "A Computer-Based Alerting System for Chemical Titles," Journal of Chemical Documentation, v. 5, n. 3 (August 1965) p. 163-165.

The Mitre Corporation
Bedford, Massachusetts

Information obtained through personal conversation during visit to company.

National Aeronautics & Space Administration
College Park, Maryland

"Automated Literature Searching Service," Journal of Chemical Documentation, v. 5, n. 2 (1965) p. 122.

"NASA Search Tape Program," Scientific Information Notes, v. 6, n. 3 (June-July 1963) p. 11.

NASA Selective Dissemination of Information Program. National Aeronautics and Space Administration, Scientific and Technical Information Division, May 21, 1964.

National Aeronautics and Space Administration. The ALPHA System: A Progress Report of Phase I of a Plan to Automate the Operation of the Redstone Scientific Information Center. Huntsville, Ala., George C. Marshal Space Flight Center, 1964. (R-COMP-A-64-1).

Wente, Van A. and Gifford A. Young. "NASA SDI Program." In: Congress of International Federation for Documentation (FID), Washington, D.C., 10-15 October 1965.

White, Herbert S. NASA/SDI Developmental Program Progress Report and Planned System. Washington, D.C., NASA Scientific and Technical Information Division, 1966.

Olin Mathieson Chemical Corporation
E.R. Squibb & Sons Division
New Brunswick, New Jersey

Freeman, Robert R., et al. "Automatic Preparation of Selected Title Lists for Current Awareness Services and as Annual Summaries

144th National Meeting of the American Chemical Society, Los Angeles, Calif., April 1963.

Sandia Corporation
Livermore Laboratory
Technical Information Division
Livermore, California

Swihart, Stanley J. and Elizabeth Bodie. An Input System for Automated Library Indexing and Information Retrieval, Including Preparation of Catalog Cards. Livermore, Calif., Sandia Corporation, 1963. (SCR-317).

U.S. Air Force Systems Command
Deputy for Foreign Technology
Wright-Patterson Air Force Base
Dayton, Ohio

Barrett, Raymond P. "CIRC - Centralized Information Reference and Control." In: Working With Semi-Automatic Documentation Systems, Warrenton, Va., AFOSR/SDC/Airlie Foundation, May 1965. p. 15-31. (AD 620-360).

Systems Development Corporation. Central Information Reference and Control (CIRC) System. Santa Monica, Calif., brochure for U.S. Air Force Systems Command, Deputy for Foreign Technology, Wright Patterson Air Force Base, Ohio, 1962.

U.S. Army Biological Laboratories
Fort Detrick, Maryland

Booz, Allen Applied Research, Inc. Mechanization Study of the Technical Information Division, U.S. Army Biological Laboratories, Fort Detrick, Maryland. Bethesda, Md. Report to Defense Supply Agency, Defense Documentation Center, under Contract DSA715 489. 1966. (BAARINC Report No. 914-1-4.)

Computer Usage Co., Inc. Selective Dissemination of Information. Bethesda, Md. Report to U.S. Army Biological Laboratories, Fort Detrick, Md., 1965. (AD 617-086).

U.S. Army Electronics Command
Information Office
Technical Information Division
Ft. Monmouth, New Jersey

Information obtained through personal conversation.

U.S. Department of the Interior
Bureau of Reclamation
Denver, Colorado

Hilf, Jack W. "Matching of Descriptors in a Selective Dissemination System." In: Automation and Scientific Communication, Washington, D.C., 26th Annual Meeting of the American Documentation Institute, 6-11 October 1963.

University of Pittsburgh
The Knowledge Availability Systems Center
Pittsburgh, Pennsylvania

Information obtained through correspondence.

APPENDIX C

BIBLIOGRAPHY PERTINENT TO THE DESIGN AND IMPLEMENTATION OF ARMY SDI SYSTEM

1. Anderson, R.R., D.R. Fitzwater, and C.R. Sage. "The Stability of Dynamic Feedback in an SDI System." In: Proceedings of the 19th International Guide Meeting, October 15, 1964.

2. "Automated Literature Searching Service," Journal of Chemical Documentation, v. 5, n. 2 (1965) p. 122.

3. Balz, Charles F. and Richard H. Stanwood. "Literature Dissemination and Retrieval Using the MERGE System." In: Automation and Scientific Communication, Washington, D.C., 26th Annual Meeting of the American Documentation Institute, 6-11 October 1963.

4. Barnes, A.B., A.A. Briggs, J. Gauss, and A. Resnick. "SDI-5, An Advanced System for Selective Dissemination of Information," ACM Proceedings of the 19th National Conference, 1964.

5. Barrett, Raymond P. "CIRC - Centralized Information Reference and Control." In: Working With Semi-Automatic Documentation Systems, Warrenton, Va., AFOSR/SDC/Airlie Foundation, May 1965. p. 15-31. (AD 620-360).

6. Belzer, Jack. "Information Retrieval on the GE-225 Computer." In: Information Retrieval in Action, Cleveland, Ohio, The Press of Western Reserve University, 1963, p. 31-40.

7. Benjamin, Curtis G. "Computers and Copyrights." Science, v. 152, n. 3719, (April 8, 1966) p. 181-184.

8. Benjamin, R., S.D. Miller, and E.S. Rowland. Selective Dissemination, (SDI). Owego, N.Y., IBM Corporation, Space Guidance Center, November 15, 1962.

9. Benjamin, R., S.D. Miller, and E.S. Rowland. Selective Dissemination, IBM 7090 Program, Owego, N.Y., IBM Corporation, Federal Systems Division, October 30, 1961.

10. Benjamin, R., S.D. Miller, and E. Scofield. Technical Report Selective Dissemination (SDI), Owego, N.Y., IBM Corporation, Federal Systems Division, October 30, 1961.

11. Booz, Allen Applied Research, Inc. Study of the Technical Information Division, U.S. Laboratories, Fort Detrick, Maryland. Bethesda, Defense Supply Agency, Defense Documentation Center. DSA715 489. 1965. (BAARINC Report No. 914-1-

12. Bottle, R.T. "A User's Assessment of Current Awareness Services," Journal of Documentation, v. 21, n. 3 (September 1965), p. 177-189.

13. Bowman, C.M. "Engineers' Personal Files and Internal Information in an Industrial Organization." Paper presented at the Symposium on Changing Aspects and New Services of Engineering Information: Fifty eighth National Meeting, Philadelphia, Pa., December 5-9, 1965, American Institute of Chemical Engineers.

14. Brandenberg, W., H.C. Fallon, C.B. Hensley, T.R. Savage, and A.J. Sowarby. Selective Dissemination of Information, SDI-2 System. Yorktown Heights, N.Y., IBM Corporation, Advanced Systems Development Division, 1961.

15. Bunnow, L.R. Study of and Proposal for a Mechanized Information Retrieval System for the Missile and Space Systems Engineering Library. Santa Monica, Calif., Douglas Aircraft Company, Inc., May 1960. (SM-37418).

16. Bushnell, R.C. Selective Dissemination of Information (SDI) Program Package for the IBM 1401. Akron, Ohio, IBM Corporation, Data Processing Division, 1962.

17. "Computer to Match Chemical Abstracts with User's Keyword Profile," Communications of the ACM, v. 9, n. 5 (May 1966) p. 401.

18. Computer Usage Co., Inc. Selective Dissemination of Information. Bethesda, Md. Report to U.S. Army Biological Laboratories, Fort Detrick, Md., 1965. (AD 617-086).

19. Croxton, Fred E. Automation Progress at RSIC: The Status of ALPHA-1. Redstone, Ala., U.S. Army Missile Command, Redstone Scientific Information Center, 1963. (RSIC-92; AD 615-813).

20. Davis, Charles J. Documentation and Information Retrieval Aspects of Army Studies: Vol. II, Annex "C" to the Army Study System. Washington, D.C., U.S. Department of the Army, 1963. (AD 602-527).

21. Davison, Roy and John Irvine. "An Announcement-And-Request System for Initial Dissemination." In: Parameters of Information Science, Volume I, Proceedings of the American Documentation Institute, 27th Annual Meeting, Philadelphia, Pa., October 5-8, 1964, p. 111-115.

22. DeVore, Charles. The Year in Navy Scientific Information. Washington, D.C., Office of Naval Research, 1965. (AD 617-315).

23. Freeman, Robert R. "Automatic Retrieval and Selective Dissemination of References from Chemical Titles: Improving the Selection Process." In: Automation and Scientific Communication, Washington, D.C., 26th Annual Meeting of the American Documentation Institute, 6-11 October 1963.

24. Freeman, Robert R., et al. "Automatic Preparation of Selected Title Lists for Current Awareness Services and as Annual Summaries," 144th National Meeting of the American Chemical Society, Los Angeles, Calif., April 1963.

25. Friedenstein, Hannah. "Alerting With Internal Abstract Bulletins," Journal of Chemical Documentation, v. 5, n. 3 (August 1965) p. 154-157.

26. Gallagher, Joan. "Experience in the Introduction of an SDI System." Paper presented at Division of Chemical Literature, American Chemical Society 151st National Meeting, Pittsburgh, Pa., March 22-24, 1966.

27. Gilbert, D.W. Army Library Today: An Inquiry Into the Character, Scope and Capabilities of Present Day Army Librarians and Library Services. 1962. 56p. (Thesis M.S. in L.S., Catnolic University of America).

28. Haas, Ann Klein. "Internal Alerting With Keyword-in-Context Indexes," Journal of Chemical Documentation, v. 5, n. 3 (August 1965) p. 160-163.

29. Halbert, Michael H. and Russell L. Ackoff. "An Operations Research Study of the Dissemination of Scientific Information." In: Proceedings of the International Conference on Scientific Information. NSF/NAS-NRC/American Documentation Institute, Washington, D.C., November 16-21, 1958.

30. Hensley, C.B. Design Developments in Information Management Through Selective Dissemination and Retrieval System. Preprints of papers presented at the 16th National Meeting of the Association for Computing Machinery, 1961.

31. Hensley, C.B. "SDI Bibliography." IBM Advanced Systems Development Division, Yorktown Heights, N.Y., September 1962.

32. Hensley, C.B. "Selective Dissemination of Information (SDI): State-of-the-Art in May, 1963." In: Spring Joint Computer Conference, Detroit, Mich., May 1963. Spartan Books, Inc., Baltimore, Md.

33. Hensley, C.B., A. Resnick, T.R. Savage, and A.J. Sowarby. Preliminary Report on the Selective Dissemination of Information System. Yorktown Heights, N.Y., IBM Corporation, Advanced Systems Development Division, 1959.

34. Hensley, C.B., T.R. Savage, A.J. Sowarby, and A. Resnick. "Selective Dissemination - Report on a Pilot Study - SDI-1 System." IBM-ASDD, Yorktown Heights, N.Y., Report 17-039, January 1961, 44p., illus.

35. Hensley, C.B., T.R. Savage, A.J. Sowarby, and A. Resnick. "Selective Dissemination of Information - A New Approach to Effective Communication." In: The Coming of Age of Information Technology, Volume VI, V. Slamecka, ed., Bethesda, Md., Documentation, Inc., 1966.

36. Hilf, Jack W. "Matching of Descriptors in a Selective Dissemination System." In: Automation and Scientific Communication, Washington, D.C., 26th Annual Meeting of the American Documentation Institute, 6-11 October 1963.

37. Hocken, Sandra. "Dissemination of Current Information," Special Libraries, v. 53, n. 2 (February 1962) p. 92-95.

38. IBM Application Program H20-0053-1, Selective Dissemination of Information System for the IBM 1401 Version I (1401-CR-01X) Manual 1: Application Description.

39. IBM Application Program H20-0054-1, Selective Dissemination of Information System for the IBM 1401 Version I (1401-CR-01X) Manual 2: User Manual.

40. IBM Application Program H20-0055-1, Selective Dissemination of Information System for the IBM 1401 Version I (1401-CR-01X) Manual 3: Operator's Manual.

41. IBM General Information Manual - Selective Dissemination of Information. San Jose, Calif., IBM Corporation, Product Publications Department, 1962.

42. "Information Project Ready to Start," Electronics Weekly, n. 232 (1965) p. 1.

43. Koriagin, G.W. Library Information Retrieval Program. Douglas Aircraft Company, Inc., Engineering Paper No. 1269, January 1962.

44. Koriagin, G.W. and L.R. Bunnow. Mechanized Information Retrieval System for Douglas Aircraft Company, Inc. Status Report. Douglas Aircraft Company, Inc., Report 39167, January 1962.

45. Kraft, Donald H. "Application of IBM Equipment to Library Mechanization, Keyword-in-Context (KWIC) Indexing and the Selective Dissemination of Information (SDI)." In: Proceedings of the 1963 Clinic on Library Applications of Data Processing, Champaign, Ill., April 28-May 1, 1963. Illini Union Bookstore, 1964, p. 133-156.

46. Kraft, Donald H. "An Operational Selective Dissemination of Information (SDI) System for Technical and Non-Technical Personnel Using Automatic Indexing Techniques." In: Automation and Scientific Communication, Washington, D.C., 26th Annual Meeting of the American Documentation Institute, 6-11 October 1963.

47. Kraft, Donald H. "Data Processing Equipment for Library Use in Clerical Tasks and Dissemination of Information." Illinois Libraries, v. 44, n. 9 (November 1962) p. 587-592.

48. Kreysa, Frank J. and W.R. Foster. "A Retrieval Profile for Current Research Information." In: Congress of International Federation for Documentation (FID), Washington, D.C., 10-15 October 1965.

49. Levy, Nathan P. and Ross M. Sigmon. "Economic Analysis of a Technical Information Dissemination System." In: Congress of International Federation for Documentation (FID), Washington, D.C., 10-15 October 1965.

50. Luhn, Hans Peter. "A Business Intelligence System," IBM Journal of Research and Development, v. 2 (October 1958) p. 314-319.

51. Luhn, Hans Peter. "Selective Dissemination of New Scientific Information with the Aid of Electronic Processing Equipment." Presented at the 126th Meeting of the American Association for the Advancement of Science, Chicago, Ill., December 29, 1959. (IBM, Yorktown Heights, N.Y., 225-1488).

52. Magnino, J.J. CIS - A Computerized Normal Text Current Awareness Technique. Yorktown Heights, N.Y., IBM Technical Information Retrieval Center, Thomas J. Watson Research Center, November 1965, Technical Report No. ITIRC-006.

53. Magnino, Joseph J., Jr. "Textual Information Retrieval and the IBM Suggestion Plan." In: Congress of International Federation for Documentation (FID), Washington, D.C., 10-15 October 1965.

54. Menzel, Herbert. "Flow of Information on Current Developments in Three Scientific Disciplines." In: Proceedings of the Federation of American Societies for Experimental Biology, v. 16 (September 1957) p. 706-711.

55. Merritt, Allen C. and Paul J. Nelson. The Engineer-Scientist and an Information Retrieval System. Yorktown Heights, N.Y., IBM Technical Information Retrieval Center, Thomas J. Watson Research Center, August 1966, Technical Report No. ITIRC-003.

56. Merrit, C. Allen. "An Operating System: The IBM Technical Information Retrieval Center." In: Working with Semi-Automatic Documentation Systems, Warrenton, Va., AFOSR/SDC/Airlie Foundation, May 1965, p. 32-39. (AD 620-360).

57. "NASA Search Tape Program," Scientific Information Notes, v. 6, n. 3 (June-July 1963) p. 11.

58. NASA Selective Dissemination of Information Program. National Aeronautics and Space Administration, Scientific and Technical Information Division, May 21, 1964.

59. National Aeronautics and Space Administration. The ALPHA System: A Progress Report of Phase I of a Plan to Automate the Operation of the Redstone Scientific Information Center. Huntsville, Ala., George C. Marshal Space Flight Center, 1964. (R-COMP-A-64-1).

60. Ofer, K.D. "SIDAR: Selective Information Dissemination and Retrieval," Journal of Chemical Documentation, v. 4, n. 1 (January 1964) p. 54-55.

61. Ossorio, Peter G. Classification Space Analysis. Boulder, Colo., University of Colorado. Report to Rome Air Development Center, under Contract AF 30(602)-3342. (RADC-TDR-64-287; AD 608-034.)

62. Ossorio, Peter G. Dissemination Research. Boulder, Colo., University of Colorado. Report to Rome Air Development Center, under Contract AF 30(602)-3342. (RADC-TR-65-314; AD 625-905).

63. Perry, J.W. Defining the Query Spectrum: The Basis for Designing and Evaluating Retrieval Methods. Unpublished paper.

64. Resnick, A. "Educational Requirements for Indexers in a Selective Dissemination System." In: Automation and Scientific Communication, Washington, D.C., 26th Annual Meting of the American Documentation Institute, 6-11 October 1963.

65. Resnick, A. "Progress Report on IBM's Selective Dissemination of Information - SDI-4 System IBM 7090 - 1401 Data Processing System." In: Automation and Scientific Communication, Washington, D.C., 26th Annual Meeting of the American Documentation Institute, 6-11 October 1963.

66. Resnick, A. "Relative Effectiveness of Document Titles and Abstracts for Determining Relevance of Documents," Science, v. 134 (1961) p. 1004-1006.

67. Resnick, A. and C.B. Hensley. "The Use of Diary and Interview Techniques in Evaluating a System for Disseminating Technical Information," American Documentation, v. 14, n. 2 (April 1963) p. 109.

68. Rice, C.N. "A Computer-Based Alerting System for Chemical Titles," Journal of Chemical Documentation, v. 5, n. 3 (August 1965) p. 163-165.

69. Sage, C.R., R.R. Anderson, and C.H. Cameron. Ames Laboratory SDI Systems Reference Manual. 1964. (U.S.A.E.C. Report IS-940).

70. Sage, C.R., R.R. Anderson, and D.R. Fitzwater. "Adaptive Information Dissemination," American Documentation, v. 16, n. 3 (July 1965) p. 185-200.

71. Sage, C.R. and D.R. Fitzwater. Ames Laboratory Selective Dissemination of Information Computer System. 1963. (U.S.A.E.C. Report IS-751).

72. Sage, C.R. and D.R. Fitzwater. "Operational Results of an Adaptive SDI System." In: Proceedings of the 19th International Guide Meeting, October 15, 1964.

73. Schulz, Louise. "RAPID, A System for Retrieval Through Automated Publication and Information Digest." In: Parameters of Information Science, Volume I. Proceedings of the American Documentation Institute, 27th Annual Meeting, Philadelphia, Pa., October 5-8, 1964, p. 79-87.

74. "Selecting the Right Data for People Who Need It," Publishers' Weekly, v. 86, n. 6 (1961) p. 64-66.

75. Sher, Irving and Eugene Garfield. "ASCA: An Automatic Current Alerting System Based on Citation Indexing." In: Congress of the International Federation for Documentation (FID), Washington, D.C., 10-15 October 1965.

76. Sowarby, A.J., et al. The SDI-3 System for the IBM 1401 Data Processing System. Yorktown Heights, N.Y., IBM Advanced Systems Development Division, 1962. (Report No. 10.3.04).

77. Spirit, Jiri and Vladimir Thannabauer. "Factors Influencing Information Dissemination in Science and Industry." In: Congress of International Federation for Documentation (FID), Washington, D.C., 10-15 October 1965.

78. Swihart, Stanley J. and Elizabeth Bodie. An Input System for Automated Library Indexing and Information Retrieval, Including Preparation of Catalog Cards. Livermore, Calif., Sandia Corporation, 1963. (SCR-317).

79. Systems Development Corporation. Central Information Reference and Control (CIRC) System. Santa Monica, Calif., brochure for U.S. Air Force Systems Command, Deputy for Foreign Technology, WPAFB, Ohio, 1962.

80. Taylor, Robert S. "The Process of Asking Questions," American Documentation, v. 13, n. 4 (October 1962) p. 391-396.

81. Tomlinson, Helen. Defining Technical Information Needs for Research Laboratory. Lackland, Tex., USAF Systems Command, Aerospace Medical Division, Personnel Research Laboratory, 1965. (Report No. PRL-TR-65-4; AD 614-820).

82. Vasarhelyi, P.E. "The Necessity of Combining Selective Dissemination of Information and Information Retrieval with Machine Translation." In: Congress of International Federation for Documentation (FID), Washington, D.C., 10-15 October 1965.

83. Warheit, I.A. "Dissemination of Information," Library Resources and Technical Services, v. 9, n. 1 (Winter 1965) p. 73-80.

84. Weik, Martin H. and Violet J. Confer. Survey of Scientific and Technical Information Retrieval Schemes Within the Department of the Army. Aberdeen Proving Ground, Md., Ballistic Research Laboratories, July 1962. (Report No. 1169).

85. Wente, Van A. and Gifford A. Young. "NASA SDI Program." In: Congress of Internation Federation for Documentation (FID), Washington, D.C., 10-15 October 1965.

86. Whaley, Fred R. and Carolyn Flanagan. "Project CADRE, Current Awareness and Document Retrieval for Engineers; A Study in Progress at Engineering Index." In: Congress of International Federation for Documentation (FID), Washington, D.C., 10-15 October 1965.

87. White, Herbert S. "The IBM DSD Technical Information Center: A Total Operating System Approach Combining Traditional Library Features and Mechanized Computer Processing." In: Automation and Scientific Communication, Washington, D.C., 26th Annual Meeting of the American Documentation Institute, 6-11 October 1963.

88. White, Herbert S. NASA/SDI Developmental Program Progress Report and Planned System. Washington, D.C., NASA Scientific and Technical Information Division, 1966.

89. Williams, A.S. "Historical Development and Present Status - Douglas Aircraft Company Computerized Library Program." In: Working With Semi-Automatic Documentation Systems, Warrenton, Va., AFOSR/SDC/Airlie Foundation, May 1965, p. 41-54. (AD 620-360).

Security Classification

DOCUMENT CONTROL DATA - R&D

(Security classification of title, body of abstract and indexing annotation must be entered when the overall report is classified)

1 ORIGINATING ACTIVITY (Corporate author) INFORMATION DYNAMICS CORPORATION 80 Main Street Reading, Massachusetts 01867	2a REPORT SECURITY CLASSIFICATION UNCLASSIFIED 2b GROUP NA

3 REPORT TITLE

Selective Dissemination of Information: Review of Selected Systems and a Design for Army Technical Libraries

4 DESCRIPTIVE NOTES (Type of report and inclusive dates)

Final Report

5 AUTHOR(S) (Last name, first name, initial)

Bivona, William A.
Goldblum, Edward J.

6 REPORT DATE August 1966	7a TOTAL NO OF PAGES 110	7b NO OF REFS
8a CONTRACT OR GRANT NO. DA 19-129-AMC-957(N) b. PROJECT NO. c d	9a ORIGINATOR'S REPORT NUMBER(S) IDC 8074 9b OTHER REPORT NO(S) (Any other numbers that may be assigned this report) ATLIS No. 8	

10 AVAILABILITY LIMITATION NOTICES

Distribution of this document is unlimited.

11 SUPPLEMENTARY NOTES This report was prepared for the Army Technical Library Improvement Studies (ATLIS) Program	12 SPONSORING MILITARY ACTIVITY U.S. Army Natick Laboratories Natick, Massachusetts

13 ABSTRACT This report presents an analysis of over eighteen small, [illegible] and large-scale systems for the selective dissemination of information (SDI). Systems are compared and analyzed with respect to design criteria and the following ten system parameters: information input, methods of indexing and abstracting, user interest profile construction, user group, profile document match strategy, profile maintenance and updating, products disseminated, processing statistics accumulated, ADP equipment utilized, and results reported. From the results of the analysis, criteria are drawn for the design of an SDI system that is generally applicable to a broad range of Army technical libraries. These design criteria are used as a basis for designing an SDI system which incorporates the following desired characteristics: broad applicability, acceptance of information in many different formats, ability to yield high relevancy for specific information requirements, and high recall for general information requirements. The system requires a minimum automatic data processing configuration and operating personnel. Operating costs are minimized by utilizing standard graphic production techniques for reproducing multiple copies of SDI notices. The user profile/document match strategy incorporates both Boolean and weighted selection criteria.

DD FORM 1 JAN 64 1473

UNCLASSIFIED

Security Classification

14 KEY WORDS	LINK A		LINK B		LINK C	
	ROLE	WT	ROLE	WT	ROLE	WT
Documentation						
Distribution						
Bibliographies						
Programming (Computers)						
Dissemination						
Information Retrieval						
Reviews						
Design						
Economics						
History						
Selection Rules						
SDI (Identifier)						
Analysis						
ATLIS Program (Identified)						

INSTRUCTIONS

1. ORIGINATING ACTIVITY: Enter the name and address of the contractor, subcontractor, grantee, Department of Defense activity or other organization *(corporate author)* issuing the report.

2a. REPORT SECURITY CLASSIFICATION: Enter the overall security classification of the report. Indicate whether "Restricted Data" is included. Marking is to be in accordance with appropriate security regulations.

2b. GROUP: Automatic downgrading is specified in DoD Directive 5200.10 and Armed Forces Industrial Manual. Enter the group number. Also, when applicable, show that optional markings have been used for Group 3 and Group 4 as authorized.

3. REPORT TITLE: Enter the complete report title in all capital letters. Titles in all cases should be unclassified. If a meaningful title cannot be selected without classification, show title classification in all capitals in parenthesis immediately following the title.

4. DESCRIPTIVE NOTES: If appropriate, enter the type of report, e.g., interim, progress, summary, annual, or final. Give the inclusive dates when a specific reporting period is covered.

5. AUTHOR(S): Enter the name(s) of author(s) as shown on or in the report. Enter last name, first name, middle initial. If military, show rank and branch of service. The name of the principal author is an absolute minimum requirement.

6. REPORT DATE: Enter the date of the report as day, month, year. If more than one date appears on the report, use date of publication.

7a. TOTAL NUMBER OF PAGES: The total page count should follow normal pagination procedures, i.e., enter the number of pages containing information.

7b. NUMBER OF REFERENCES: Enter the total number of references cited in the report.

8a. CONTRACT OR GRANT NUMBER: If appropriate, enter the applicable number of the contract or grant under which the report was written.

8b, 8c, & 8d. PROJECT NUMBER: Enter the appropriate military department identification, such as project number, subproject number, system numbers, task number, etc.

9a. ORIGINATOR'S REPORT NUMBER(S): Enter the official report number by which the document will be identified and controlled by the originating activity. This number must be unique to this report.

9b. OTHER REPORT NUMBER(S): If the report has been assigned any other report numbers *(either by the originator or by the sponsor)*, also enter this number(s).

10. AVAILABILITY/LIMITATION NOTICES: Enter any limitations on further dissemination of the report, other than those imposed by security classification, using standard statements such as:

(1) "Qualified requesters may obtain copies of this report from DDC."

(2) "Foreign announcement and dissemination of this report by DDC is not authorized."

(3) "U. S. Government agencies may obtain copies of this report directly from DDC. Other qualified DDC users shall request through _____."

(4) "U. S. military agencies may obtain copies of this report directly from DDC. Other qualified users shall request through _____."

(5) "All distribution of this report is controlled. Qualified DDC users shall request through _____."

If the report has been furnished to the Office of Technical Services, Department of Commerce, for sale to the public, indicate this fact and enter the price, if known.

11. SUPPLEMENTARY NOTES: Use for additional explanatory notes.

12. SPONSORING MILITARY ACTIVITY: Enter the name of the departmental project office or laboratory sponsoring *(paying for)* the research and development. Include address.

13. ABSTRACT: Enter an abstract giving a brief and factual summary of the document indicative of the report, even though it may also appear elsewhere in the body of the technical report. If additional space is required, a continuation sheet shall be attached.

It is highly desirable that the abstract of classified reports be unclassified. Each paragraph of the abstract shall end with an indication of the military security classification of the information in the paragraph, represented as *(TS)*, *(S)*, *(C)*, or *(U)*.

There is no limitation on the length of the abstract. However, the suggested length is from 150 to 225 words.

14. KEY WORDS: Key words are technically meaningful terms or short phrases that characterize a report and may be used as index entries for cataloging the report. Key words must be selected so that no security classification is required. Identifiers, such as equipment model designation, trade name, military project code name, geographic location, may be used as key words but will be followed by an indication of technical context. The assignment of links, rules, and weights is optional.